

## **Lincoln University Digital Thesis**

### **Copyright Statement**

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- you will use the copy only for the purposes of research or private study
- you will recognise the author's right to be identified as the author of the thesis and due acknowledgement will be made to the author where appropriate
- you will obtain the author's permission before publishing any material from the thesis.

# The quid pro quo of dissolved organic matter biodegradability in agricultural soils: investigations of quantitative and biochemical aspects

---

A thesis submitted in partial fulfilment  
of the requirements for the degree of

Doctor of Philosophy  
at  
Lincoln University

By  
Tihana Vujinović

---

Lincoln University  
2019

### **Published parts of this Thesis**

Vujinović, T., Clough, T.J., Curtin, D., Meenken, E.D., Lehto, N.J. and Beare, M.H. (2019). Quantity and biodegradability of dissolved organic matter released from sequentially leached soils, as influenced by the extent of soil drying prior to rewetting. *Soil Research*, 57(4), 374–386.

*I dedicate this piece of work, and with it all my achievements  
and who am I today, to my family*

*Ljubica, a life warrior*

*Aleksandar, truth of emotions*

*Marija, spirit of life*

*Radovan, joy*

*Franco, window of soul*

*Sandra, beauty*

*Nikola, dedication*



# Abstract

*Dissolved Organic Matter* (DOM) serves as a primary source of mineralizable carbon (C), nitrogen (N), and other macroelements for plants and soil microbiota, acting as a linchpin of nutrient dynamics and energy flows through ecosystems. However, its mineralization through microbial activity can potentially lead to serious environmental threats. These include inflated losses as carbon dioxide (CO<sub>2</sub>) into the atmosphere, contributing to the global greenhouse effect, or enhanced groundwater contamination and reduced soil fertility, due to leaching. The balance between processes of formation, utilization and preservation of DOM is a function of a wide range of biological, environmental and anthropogenic factors. As a result, current understanding of the quantity, composition and biodegradability potential of this dynamic soil organic matter (SOM) component in rather complex and changing systems, such as agricultural soils, remains moot.

This study attempted to shed light on the quantitative and biochemical aspects of DOM underpinning its biodegradability potential. What are the features that best describe the mineralization response? How do changes in soil moisture and land use affect these and are there preferred forms and pathways that the microbial community exploits? These are some of the spread puzzle pieces of DOM studies that the present study aimed to combine.

The first part of the study investigated the effect of different soil moisture levels on the quantitative and qualitative changes of DOM released from sequentially leached arable and grassland soils. The extent of soil drying before rewetting had a strong impact on the quantity of DOM released with sequential leaching, but these changes were inconsistent between land uses and poorly related to DOM biodegradability. Advancing drought levels progressively decreased differences in DOM quality, assessed by means of specific ultraviolet (UV) absorbance at 254 nm, but there remained notable changes in DOM released over the leaching course, indicating an increasing proportion of water-soluble and relatively biodegradable aromatic and/or conjugated structures. These proportions and their responsiveness to biodegradation varied, however, between land uses, questioning the involvement of components of different origin and functionality. A soil with three contrasting management histories was further considered, with the aim of identifying and characterizing, by means of fluorescence and UV indicators, specific fractions that might

explain DOM mineralization response to land use. Anion exchange resin treatment of cold (CW) and hot water (HW) extracts revealed an uncharged/protonated, aliphatic, likely microbially derived sub-fraction, with contrasting biodegradability in CW and HW extracts, as opposed to the negatively charged and/or hydrophobic, aromatic, plant and SOM-derived fraction retained by the resin, which had comparable biodegradability between CW and HW extracts and land uses. The Random Forest algorithm inferred the degree of humification to be an important aspect affecting the mineralization of the CW fraction, but not of the HW fraction, conversely characterized by a large amount of phenols and organic N content. In contrast, DOM concentration was not a factor limiting biodegradability.

Open questions as to what caused these inconsistencies in the mineralization response among DOM fractions and sub-fractions were finally approached at the molecular level in the last part of the study, combining data mining of fluorescence Excitation-Emission Matrices with high resolution Liquid Chromatography–Mass Spectrometry. Small peptides and quaternary ammonium compounds were attributed a leading role in DOM mineralization, which was, in contrast, demoted by optically active, less oxygenated, lipid-like and phenolic compounds. This study confirms the importance of compositional aspects of DOM in shaping the microbial metabolism and the fate of DOM biodegradation products.

**Keywords:** Anion exchange, Biodegradation, Data mining, Dissolved organic matter, Drying–rewetting, Fluorescence spectroscopy, Land use, Liquid chromatography–mass spectrometry, Parallel factor analysis, Principal component analysis, Random forest, Ultraviolet spectroscopy, van Krevelen diagrams, Water extractable organic matter.

# Acknowledgments

At the end of this journey, I would like to thank my main supervisor, Prof. Tim Clough, for recognizing my motivation to challenge and be challenged by soil organic matter studies and for giving me the opportunity to do and complete my research here in New Zealand, also supporting my successful scholarship application. I would also thank him for being patient with me, for teaching me how to keep focused on the aim in order to draw my cartoon, and reminding me that “good research brings few answers and lots of questions”.

This work wouldn't have been possible without the joint collaboration of Lincoln University with Plant and Food Research and my associate supervisor, Dr. Mike Beare, who provided great financial support and the laboratory equipment to make this work happen. I would also like to thank Mike for opening the door of the Plant and Food family for me, which allowed me to have a second home and meet many good people, share experience and learn from them. I would also thank him and Paula for hosting me in their lovely home. I want to thank my co-supervisors, Dr. Denis Curtin, for his care and interest in me as a person and researcher and the many coffees and scientific debates, and Dr. Nik Lehto, for his interest in my research and the opportunity to collaborate with the University of Waikato. I would like to thank my supervisors for their suggestions and revisions, but also for preparing me for future challenges.

A big thanks to Prof. Jim Moir, for his great support and encouragement. Thanks also to Prof. Gianluca Gorni, who assisted me with the formatting of this thesis, regardless of time and space divergence. I would also thank all the laboratory staff at PFR, in particular Chris Dunlop, Michelle Peterson, Kathryn Lehto and Weiwen Qiu, the data science group, and Nigel Joyce for his assistance with the mass spectrometry. I am thankful to Dr. Adam Hartland and Peter Reutemann (University of Waikato) for supporting my work with fluorescence spectroscopy and data mining. I would then like to thank my former Professor and friend, Maria De Nobili, for always believing in me, being there for me and encouraging me to pursue my doctoral research.

The dedicated space for these acknowledgments is not enough to thank all the people that supported me throughout this journey. Thanks to my officemates, David and Camilla, for creating a family workspace for us, for all the smiles and

tears we shared, and special thanks to David, for his endless care, patience and support in any time, and all the amazingly thoughtful goodie boxes. Thanks then to my PFR officemate Carmen, who has shared with me everyday challenges, but also joy and cozy moments, the ones we like. A big thanks to all my colleagues and ‘New Zealand’ friends, especially Monica, Amy, Dirk, Alvand, Carolina, Kritarth, Dharini, Renato, Yuki, Léonie, Lingying and Jun, for their understanding and support, especially throughout the hardest periods of this story. Special thanks to Zach Simpson, for offering me his shoulder and skills in the moment of need and being patient with me, which I will never forget. Another big mention to my PFR colleague Jian (Frank) Liu, for his big technical and moral support. A big thank you Peter Carey, for all the care, the comfort food supplies and for every thought dedicated to cheer me up, e.g. racing with your cars. Roger Cresswell, you know the silver lining in a cloudy day? Summing up all days like this, I was able to have many sunny days: thank you. And thanks to our department mom Amal Torky, for her unique warmth and encouragement.

A great thanks to Craig and Rebekah Tregurtha, for their proactive and positive attitude, their prayers and help in the moments of need. Thanks to Rob and Ann Blackburne, for their empathy and help, and to Jayne Smith, for her ear and support throughout the very important final stage of my PhD. A great thanks to Myra Duthie for her administrative support and care.

I would then like to thank all my friends back home and all over the world, who have always supported me and have been there for me, despite the distance, especially Ines, Martina, Francesco, Mauricio, Luca, Paolo, Vito, Andrea, Martin, Petra, Namira, Igor and Lina. Elena, thank you for always believing in me and sending me all your care and support. Thank you Suzy, for being like a sister for me. A special mention to Vincenzo De Caria, for his long and true friendship, and for his technical support to get this thesis finally compiled. Loris, you fueled every single day of these years of distance, and showed me the reason to fight for; I cherish the day we met. Thanks also to all those that I didn’t mention, but that in a way or another enriched this experience, while also making it harder for me, because they were preparing me for something big.

Most importantly, I want to thank my family, for their unconditional love and support, my grandma Marija for her incredible heart, affection and patience, and especially my mother Ljubica, who has been through every single moment with me, and my life’s pillar.

It was certainly a big story.

# Contents

<b>Abstract</b>	<b>iv</b>
<b>Acknowledgments</b>	<b>vi</b>
<b>List of Figures</b>	<b>xii</b>
<b>List of Tables</b>	<b>xviii</b>
<b>Abbreviations</b>	<b>xx</b>
<b>1 Introduction</b>	<b>1</b>
1.1 General introduction and scope . . . . .	1
1.1.1 Background . . . . .	1
1.1.2 Research objectives . . . . .	2
1.1.3 Thesis structure . . . . .	3
1.2 Literature review . . . . .	4
1.2.1 DOM in natural and agricultural ecosystems . . . . .	4
1.2.2 Factors and sources of DOM release . . . . .	6
1.2.3 The relationship between DOM mineralization and its properties: bioavailability and biodegradability . . . . .	11
<b>2 Materials and Methods</b>	<b>15</b>
2.1 General chemical analyses . . . . .	15
2.1.1 Elemental analyses: total carbon and nitrogen . . . . .	15
2.1.2 Dissolved organic carbon . . . . .	16
2.1.3 Dissolved nitrogen . . . . .	16
2.1.4 pH . . . . .	17
2.1.5 Cation exchange capacity . . . . .	17
2.2 Spectroscopic techniques . . . . .	18
2.2.1 Ultraviolet absorbance . . . . .	18
2.2.2 Fluorescence . . . . .	19
2.3 Chromatographic techniques . . . . .	24

2.3.1	Anion exchange chromatography . . . . .	24
2.3.2	Liquid chromatography . . . . .	25
2.4	Mass spectrometry . . . . .	27
2.4.1	High-resolution accurate-mass (HRAM): Orbitrap . . . . .	27
2.4.2	Detection of unknown compounds . . . . .	30
2.4.3	van Krevelen diagrams and structural determinants . . . . .	31
2.5	Statistical analyses and chemometric techniques . . . . .	32
2.5.1	Analysis of variance . . . . .	32
2.5.2	Restricted maximum likelihood and linear mixed models . . . . .	34
2.5.3	Multivariate analyses . . . . .	36
2.5.4	Machine learning techniques . . . . .	42
<b>I</b>	<b>Behind the scenes of DOM biodegradability</b>	<b>45</b>
<b>3</b>	<b>Quantity and biodegradability of DOM released from sequentially leached soils, as influenced by the extent of soil drying prior to rewetting</b>	<b>47</b>
3.1	Introduction . . . . .	48
3.2	Materials and methods . . . . .	50
3.2.1	Soil sampling and characterization . . . . .	50
3.2.2	Experimental treatments . . . . .	51
3.2.3	Leachate collection . . . . .	52
3.2.4	Leachate sample analyses and characterization . . . . .	53
3.2.5	Water extractable organic C (WEOC) . . . . .	53
3.2.6	Biodegradation assay . . . . .	54
3.2.7	Statistical analysis . . . . .	54
3.3	Results and discussion . . . . .	55
3.3.1	Soil properties . . . . .	55
3.3.2	DOC in soil leachates . . . . .	55
3.3.3	Leachate N . . . . .	59
3.3.4	SUVA <sub>254</sub> . . . . .	61
3.3.5	Leachate pH . . . . .	67
3.3.6	Biodegradability of DOC . . . . .	67
3.4	Conclusions . . . . .	71
<b>4</b>	<b>Quantitative and qualitative properties of WEOM fractions in relation to biodegradability: an anion exchange fractionation study</b>	<b>73</b>
4.1	Biodegradability and qualitative characteristics of WEOM fractions in a soil with different land use histories . . . . .	75
4.1.1	Introduction . . . . .	75

4.1.2	Materials and methods . . . . .	79
4.1.3	Results . . . . .	84
4.1.4	Discussion . . . . .	92
4.1.5	Conclusions . . . . .	101
4.2	The effect of pH on the retention efficacy of WEOM by anion exchange resin . . . . .	102
4.2.1	Rationale . . . . .	102
4.2.2	Experimental . . . . .	103
4.2.3	Results and discussion . . . . .	105
4.2.4	Conclusions . . . . .	108

## II Molecular aspects of WEOM fractions in relation to their biodegradability 109

5	Searching for proxies of WEOM biodegradability: an integrated Fluorescence Data Mining and Liquid Chromatography–Mass Spectrometry approach . . . . .	111
5.1	Introduction . . . . .	112
5.2	Materials and methods . . . . .	116
5.2.1	Soils, extracts and WEOM fractionation . . . . .	116
5.2.2	WEOM biodegradability assay . . . . .	117
5.2.3	WEOM characterization . . . . .	117
5.2.4	Suwannee River humic acid and peat extracts . . . . .	118
5.2.5	Fluorescence analyses . . . . .	118
5.2.6	Preliminary parallel factor analysis . . . . .	118
5.2.7	Data mining . . . . .	119
5.2.8	LC–MS analysis . . . . .	120
5.2.9	LC–MS data processing . . . . .	121
5.2.10	PCA . . . . .	123
5.3	Results . . . . .	124
5.3.1	WEOM fingerprinting . . . . .	124
5.3.2	Algorithm evaluation and selection of PARAFAC components . . . . .	124
5.3.3	Preliminary filtering evaluation . . . . .	126
5.3.4	Biochemical profile of resolved compounds . . . . .	127
5.3.5	Changes in molecular characteristics of WEOM . . . . .	129
5.3.6	Optical and molecular aspects in relation to WEOM biodegrad- ability . . . . .	132
5.4	Discussion . . . . .	137
5.4.1	RF identification of PARAFAC components . . . . .	137
5.4.2	Analysis of EEMs and PARAFAC components . . . . .	137

---

5.4.3	Current understanding of van Krevelen diagrams . . . . .	139
5.4.4	Peptides or nonpeptides? . . . . .	143
5.4.5	A closer look at the character of “not matched” compounds . . . . .	146
5.4.6	Inconsistent aspects of structural determinants . . . . .	148
5.4.7	Relationship with WEOM biodegradability . . . . .	149
5.5	Conclusions . . . . .	152
<b>6</b>	<b>Corollaries and Implications for future work</b>	<b>155</b>
<b>Appendices</b>		
<b>A</b>		<b>161</b>
<b>B</b>		<b>167</b>
B.1	Anion exchange resin equilibration times . . . . .	175
B.2	Selection and calculation of the humification index . . . . .	177
<b>C</b>		<b>183</b>
C.1	PARAFAC data mining . . . . .	189
<b>Bibliography</b>		<b>241</b>



# List of Figures

1.1	Conceptual model of the C mineralization process (modified from Curtin et al., 2012) . . . . .	6
2.1	The Jablonski diagram . . . . .	22
2.2	The Orbitrap mass analyzer . . . . .	29
2.3	An example of residuals' plots from the LMM fitted with REML . .	36
2.4	Eigenvectors of a multivariate Gaussian distribution . . . . .	37
2.5	PARAFAC decomposition of the trilinear EEM structure in five components . . . . .	40
3.1	Boxplots showing total dissolved organic C (sum of DOC in four pore volumes) leached from an arable ( <i>a</i> ) and a grassland soil ( <i>b</i> ) at varying degrees of $\theta_g$ : field capacity (FC), 15%, 8% and <2% $\theta_g$ (air dry; AD) . . . . .	57
3.2	The relationship between total dissolved organic C (DOC; mg g <sup>-1</sup> soil C) of leachates and extracts from an arable ( <i>a</i> ) and a grassland soil ( <i>b</i> ) at FC, 15% $\theta_g$ , 8% $\theta_g$ and air dry (AD) . . . . .	58
3.3	Dissolved organic C content (DOC; mean $\pm$ s.e.m.) in the pore volumes leached from an arable ( <i>a</i> ) and a grassland soil ( <i>b</i> ) at four levels of $\theta_g$ . . . . .	60
3.4	Dissolved organic N content (DON; mean $\pm$ s.e.m.) in the pore volumes leached from a grassland soil at four levels of $\theta_g$ . . . . .	61
3.5	The relationship between the concentration of dissolved organic C (DOC) and SUVA <sub>254</sub> in leachates of different $\theta_g$ treatments of an arable ( <i>a</i> ) and a grassland soil ( <i>b</i> ) . . . . .	64
3.6	The pH of the different pore volumes leached from an arable ( <i>a</i> ) and a grassland soil ( <i>b</i> ) at their designated $\theta_g$ before rewetting . . . . .	66
3.7	Cumulative mineralization of dissolved organic C (DOC-C <sub>min</sub> ) in the different pore volumes leached from an arable ( <i>a</i> ) and a grassland soil ( <i>b</i> ) at their designated $\theta_g$ before rewetting . . . . .	69

3.8	Cumulative mineralization of dissolved organic C (DOC- $C_{min}$ ) as a function of DOC concentration ( $a$ ) and SUVA <sub>254</sub> ( $b$ ) in leachates of a grassland soil at different $\theta_g$ before rewetting . . . . .	70
4.1	Cumulative WEOC mineralized (WEOC <sub>min</sub> ) over 14 days in non- and post-resin treated cold (CW) and hot water (HW) extracts of a soil under different land uses . . . . .	92
4.2	Outputs of the Random Forest algorithm describing the variable importance in explaining the WEOC mineralization (WEOC <sub>min</sub> ) response: a) Increased Mean square-error (%IncMSE) and b) Increased Node Impurity Index (IncNodePurity) . . . . .	93
4.3	Concentrations of WEOC and WEON in cold (CW) and hot water (HW) extracts at their original or adjusted value of pH before (resin:extract ratio = 0) and after the resin treatment (resin:extract ratio = 0.05 (1:20), 0.075 (1.5:20) and 0.1 (2:20)) . . . . .	104
4.4	Changes in pH values of cold (CW) and how water (HW) extracts following the resin treatment ( $n = 3$ ); resin:extract ratio = 0 (non-resin treated), 0.05 (1:20), 0.075 (1.5:20) and 0.1 (2:20) . . . . .	106
4.5	Changes in SUVA <sub>254</sub> and $a_{254}$ values of cold (CW) and how water (HW) extracts following the resin treatment ( $n = 3$ ); resin:extract ratio = 0 (non-resin treated), 0.05 (1:20), 0.075 (1.5:20) and 0.1 (2:20) 107	
5.1	Distribution of compound classes designated according to Rivas-Ubach et al. (2018) in pre- and post-incubation, non- and post-resin treated, cold (CW) and hot water (HW) extracts of different land uses 130	
5.2	Distribution of compounds not matching any of the classes defined by Rivas-Ubach et al. (2018) across six discrete categories based on $AI_{mod}$ (Koch and Dittmar, 2006), H/C and O/C ratios, and heteroatom contents in pre- and post-incubation, non- and post-resin treated, cold (CW) and hot water (HW) extracts of different land uses 131	
5.3	van Krevelen diagrams of non- and post-resin treated, cold (CW) and hot water (HW) extracts before and after the 14-day incubation, plotted against DBE-O and relative to their chromatography/ionization mode of detection . . . . .	133
5.4	van Krevelen diagrams of non- and post-resin treated, cold (CW) and hot water (HW) extracts before and after the 14-day incubation, plotted against the number of nitrogen atoms (# N) and relative to their chromatography/ionization mode of detection . . . . .	134

5.5	Principal component analysis of PARAFAC component scores (C1–C5), weighted molecular variables, WEOM optical properties and biodegradability indicators. Axes were rotated to visualize the vector loadings in the second dimension (PC2) . . . . .	136
5.6	van Krevelen diagram of all compounds revealed in the pre-incubation, non-resin treated hot water extracts across different land uses . . . .	140
A.1	Experimental drying curves of an arable and a grassland soil utilized for the establishment of the designated $\theta g$ treatments at low temperature (4–6°C) and constant humidity . . . . .	162
A.2	UV absorbance at 254 nm of the pore volumes leached from an arable (a) and a grassland soil (b) at their designated $\theta g$ . . . . .	163
A.3	Arable site . . . . .	164
A.4	Grassland site . . . . .	164
A.5	Leachate collection from repacked arable and grassland soils . . . .	165
A.6	Infrared gas analyzer (IRGA) setup and gas sampling . . . . .	165
B.1	Anion exchange stoichiometry of WEOC for chloride in cold (CW) and hot water (HW) extracts . . . . .	168
B.2	Anion exchange stoichiometry of total anions for chloride in cold (CW) and hot water (HW) extracts . . . . .	169
B.3	Resin removal isotherms of cold (CW) and hot water (HW) extracts showing the equilibrium adsorption capacity ( $q_e$ ), estimated as the amount of adsorbed WEOC (mg) per unit mass resin (g), associated with the non-removable WEOC sub-fraction . . . . .	170
B.4	Boxplots showing the difference in mineral N (Min N = $\text{NO}_3^- + \text{NH}_4^+$ ) before and after the 14-day incubation of non-resin treated and post-resin treated cold (CW) and hot water (HW) extracts of a soil under different land uses . . . . .	171
B.5	Random Forest regression analysis revealing the relationship between the $A_{465}$ Milori Humification Index and $\text{WEOC}_{min}$ in non- and post-resin treated cold (CW) and hot water (HW) extracts . . . . .	172
B.6	Relationship between utilized WEOC ( $\Delta\text{WEOC}$ ) and WEON ( $\Delta\text{WEON}$ ) after the 14-day incubation of cold (CW) and hot (HW) extracts . .	173
B.7	Relationship between the absorption coefficient at 254 nm ( $a_{254}$ ) and mineral N (Min N = $\text{NO}_3^- + \text{NH}_4^+$ ) for cold (CW) and hot (HW) extracts . . . . .	173
B.8	Millenium Tillage Trial split-plot design with three replications for each land use treatment . . . . .	174
B.9	Determination of resin equilibration times and maximal WEOC removal from cold (CW) and hot water (HW) extracts . . . . .	175

B.10	Properties of the anion exchange resin used in this study . . . . .	176
B.11	Smoother fits and actual values for the $A_{465}$ Milori Humification Index in non- and post-resin treated cold water extracts (CWE) . .	178
B.12	Smoother fits and actual values for the $A_{465}$ Milori Humification Index in non- and post-resin treated hot water extracts (HWE) . .	179
B.13	Smoother fits and actual values for the $A_{465}$ Milori Humification Index in non- and post-resin treated cold water extracts (CWE) after the 14-day incubation . . . . .	180
B.14	Smoother fits and actual values for the $A_{465}$ Milori Humification Index in non- and post-resin treated hot water extracts (HWE) after the 14-day incubation . . . . .	181
C.1	Variables included in the data mining approach, showing pre- and post-incubation changes in non- and post-resin treated cold (CW) and hot water (HW) extracts of a soil under different land uses (Fig. <i>a-d</i> ). Phenols (Fig. <i>e</i> ) were not detected in post-resin treated samples	183
C.2	EEMs of WEOM in non- and post-resin treated cold water (CW) extracts of a soil under three different land uses before (pre-) and after (post-) the 14-day incubation . . . . .	184
C.3	EEMs of WEOM in non- and post-resin treated hot water (HW) extracts of a soil under three different land uses before (pre-) and after (post-) the 14-day incubation . . . . .	185
C.4	Pre- and post-incubation changes in WEOM fluorescence of non- and post-resin treated cold water (CW) extracts of a soil under three different land uses . . . . .	186
C.5	Pre- and post-incubation changes in WEOM fluorescence of non- and post-resin treated hot water (HW) extracts of a soil under three different land uses . . . . .	187
C.6	EEM of the post-incubation phenolic mixture . . . . .	188
C.7	Fluorescence emission at 330 nm excitation with varying chloride ( $\text{Cl}^-$ ) concentrations before and after the resin treatment . . . . .	188
C.8	Root mean square error (RMSE) of the RF algorithm in estimating WEOC mineralization ( $\text{WEOC}_{min}$ ) and mineral N (Min N) production over 14 days and WEOM pre- and post-incubation optical properties ( $\text{SUVA}_{254}$ and $a_{254}$ ) using PARAFAC components . . . .	190
C.9	Root mean square error (RMSE) of the RF algorithm in estimating WEOC and WEON utilization ( $\Delta\text{WEOC}$ and $\Delta\text{WEON}$ ) over 14 days, and pre- and post-incubation WEOC, WEON and phenols concentrations using PARAFAC components . . . . .	191

C.10	Root mean square error (RMSE) of the LWL-GPD algorithm in estimating WEOC mineralization ( $WEOC_{min}$ ) and mineral N (Min N) production over 14 days and WEOM pre- and post-incubation optical properties ( $SUVA_{254}$ and $a_{254}$ ) using PARAFAC components	192
C.11	Root mean square error (RMSE) of the LWL-GPD algorithm in estimating WEOC and WEON utilization ( $\Delta WEOC$ and $\Delta WEON$ ) over 14 days, and pre- and post-incubation WEOC, WEON and phenols concentrations using PARAFAC components	193
C.12	Actual vs. predicted patterns of WEOC mineralization ( $WEOC_{min}$ ), mineral N (Min N) production, pre- and post-incubation WEOM optical properties ( $SUVA_{254}$ and $a_{254}$ ) and phenolic content determined by six (blue dots) or nine (red dots) PARAFAC components using the RF algorithm	194
C.13	Actual vs. predicted patterns of WEOC and WEON utilization ( $\Delta WEOC$ and $\Delta WEON$ ) and pre- and post-incubation WEOC and WEON concentrations determined by six (blue dots) or nine (red dots) PARAFAC components using the RF algorithm	195
C.14	Actual vs. predicted patterns of WEOC mineralization ( $WEOC_{min}$ ), mineral N (Min N) production, pre- and post-incubation WEOM optical properties ( $SUVA_{254}$ and $a_{254}$ ) and phenolic content determined by six (blue dots) or nine (red dots) PARAFAC components using the LWL-GPD algorithm	196
C.15	Actual vs. predicted patterns of WEOC and WEON utilization ( $\Delta WEOC$ and $\Delta WEON$ ) and pre- and post-incubation WEOC and WEON concentrations determined by six (blue dots) or nine (red dots) PARAFAC components using the LWL-GPD algorithm	197
C.16	van Krevelen diagrams of non- and post-resin treated, cold (CW) and hot water (HW) extracts before and after the 14-day incubation, plotted against DBE/C and relative to their chromatography/ionization mode of detection	198
C.17	van Krevelen diagrams of non- and post-resin treated, cold (CW) and hot water (HW) extracts before and after the 14-day incubation, plotted against the molecular weight (MW) and relative to their chromatography/ionization mode of detection	199
C.18	van Krevelen diagrams of compounds classified according to Rivas-Ubach et al. (2018) in non- and post-resin treated, cold water (CW) extracts before and after the 14-day incubation, relative to their chromatography/ionization mode of detection and presence or absence of nitrogen in their structure	200

C.19	van Krevelen diagrams of compounds classified according to Rivas-Ubach et al. (2018) in non- and post-resin treated, hot water (HW) extracts before and after the 14-day incubation, relative to their chromatography/ionization mode of detection and presence or absence of nitrogen in their structure . . . . .	201
C.20	Principal component analysis loadings of PARAFAC component scores (C1–C5), weighted molecular determinants, WEOM optical properties and biodegradability indicators for PC1 and PC2 . . . . .	206
C.21	van Krevelen diagrams of the peat and Suwannee River humic acid extracts plotted against the molecular weight (MW) and relative to their chromatography/ionization mode of detection . . . . .	213
C.22	van Krevelen diagrams of the peat and Suwannee River humic acid extracts plotted against the number of nitrogen atoms (# N) and relative to their chromatography/ionization mode of detection . . . . .	214
C.23	van Krevelen diagrams of the peat and Suwannee River humic acid extracts plotted against DBE-O and relative to their chromatography/ionization mode of detection . . . . .	215
C.24	van Krevelen diagrams of the peat and Suwannee River humic acid extracts plotted against DBE/C and relative to their chromatography/ionization mode of detection . . . . .	216
C.25	van Krevelen diagrams of compounds classified according to Rivas-Ubach et al. (2018) in the peat and Suwannee River humic acid extracts plotted relative to their chromatography/ionization mode of detection . . . . .	217
C.26	Principal component analysis of all PARAFAC component scores (C1–C6), fluorescence peaks, maximum intensities ( $I_{max}$ ) and relative excitation (Ex) and emission (Em) wavelengths, and all quantitative and qualitative variables investigated in Chapter 4 . . . . .	218
C.27	LC–MS processing workflow . . . . .	219

# List of Tables

3.1	Physical and chemical properties of arable and grassland soils. . . .	51
3.2	Physical properties of repacked arable and grassland soil cores. . . .	52
3.3	SUVA <sub>254</sub> values (L mg <sup>-1</sup> m <sup>-1</sup> ) in the pore volumes leached from an arable and a grassland soil at their $\theta_g$ treatment before rewetting. .	62
4.1	General properties of surface soils (0–7.5 cm) kept under a pasture, a no-tillage or a fallow regime for 13 years. . . . .	85
4.2	Absolute amounts of WEOC and WEON released following cold (CW) and hot water (HW) extractions of soils from different land use treatments. . . . .	86
4.3	Quantitative and qualitative properties of non- and post-resin treated cold (CW) and hot water (HW) extracts, as affected by the land use.	87
4.4	REML output indicating the statistical significance of land use, extraction method, resin treatment and their interaction on the quantitative and qualitative variables investigated. . . . .	88
5.1	Range of maximum fluorescence intensities ( $I_{max}$ ) and relative excitation (Ex) and emission (Em) wavelengths of non- and post-resin treated, cold (CW) and hot water (HW) extracts across three land uses ( $n = 9$ for each land use), before and after the 14-day incubation.	125
5.2	Summary statistics for the Random Forest regression between actual and estimated variables using six PARAFAC components. . . . .	126
5.3	Initial median values of coefficients of variation (CV) of all peak areas relative to their chromatography/ionization mode, and peak area distribution of only compounds with assigned molecular formulae, following CV and heteroatom/C filtering. The median peak area value laid around the noise threshold and was therefore selected as a filtering criteria for each batch. The number of compounds prior to each filtering step is included in brackets. . . . .	127

5.4	Number of molecules and relative proportion of designated biochemical classes across the full data set. Counts refer to unique observations with the highest intensity area across chromatographic separations and ionization modes. . . . .	128
A.1	Multivariate linear regression for SUVA <sub>254</sub> as a function of DOC concentration (Fig. 3.5) in leachates from an arable and a grassland soil at different $\theta g$ before rewetting. . . . .	161
A.2	Multivariate linear regression for DOC biodegradability (DOC- $C_{min}$ ) as a function of DOC concentration (Fig. 3.8a) in leachates from a grassland soil at different $\theta g$ before rewetting. . . . .	162
B.1	Concentration of anions in cold (CW) and hot water (HW) extracts before (a) and after (b) the resin treatment. Values are means ( $n = 3$ ); s.e.d = standard error of differences of means; l.s.d. = least significant difference ( $\alpha = 0.05$ ). . . . .	167
C.1	Summary statistics for the RF regression between actual and estimated WEOC and WEON utilization ( $\Delta$ WEOC and $\Delta$ WEON), and pre- and post-incubation WEOC and WEON using six PARAFAC components. . . . .	189
C.2	Mean scores of PARAFAC components (C1–C5) and weighted average of molecular determinants of pre-incubation, non-resin treated cold (CW) and hot water (HW) extracts of different land uses utilized for PCA analysis. . . . .	202
C.3	Mean scores of PARAFAC components (C1–C5) and weighted average of molecular determinants of pre-incubation, post-resin treated cold (CW) and hot water (HW) extracts of different land uses utilized for PCA analysis. . . . .	203
C.4	Mean scores of PARAFAC components (C1–C5) and weighted average of molecular determinants of post-incubation, non-resin treated cold (CW) and hot water (HW) extracts of different land uses utilized for PCA analysis. . . . .	204
C.5	Mean scores of PARAFAC components (C1–C5) and weighted average of molecular determinants of post-incubation, post-resin treated cold (CW) and hot water (HW) extracts of different land uses utilized for PCA analysis. . . . .	205
C.6	List of compounds discussed in Chapter 5, including the molecular formulae, monoisotopic masses, structures and chromatography/ionization modes of detection. . . . .	207



# Abbreviations

$\beta:\alpha$  Freshness Index.

$\theta_g$  Gravimetric Soil Water Content.

$a_{254}$  Absorption Coefficient at 254 nm.

$A_{465}$  Milori Humification Index.

**AER** Anion Exchange Resin.

**AI** Aromaticity Index.

**ANOVA** Analysis of Variance.

**CW** Cold Water.

**DBE** Double Bond Equivalent.

**DOC** Dissolved Organic Carbon.

**DOC- $C_{min}$**  Dissolved Organic Carbon Mineralization.

**DOM** Dissolved Organic Matter.

**DON** Dissolved Organic Nitrogen.

**EEMs** Excitation-Emission Matrices.

**FI** Fluorescence Index.

**FWHM** Full Width at Half Maximum.

**GPD** Gaussian Processes.

**HILIC** Hydrophilic Interaction Liquid Chromatography.

**HIX** Humification Index.

**HRAM** High-Resolution Accurate-Mass.

**HW** Hot Water.

**IEC** Ion Exchange Chromatography.

**LC–MS** Liquid Chromatography–Mass Spectrometry.

**LMM** Linear Mixed Model.

**LWL** Locally Weighted Learning.

**Min N** Mineral Nitrogen.

**PARAFAC** Parallel Factor Analysis.

**PCA** Principal Component Analysis.

**PV** Pore Volume.

**REML** Restricted Maximum Likelihood.

**RF** Random Forest.

**RMSE** Root Mean Square Error.

**SUVA<sub>254</sub>** Specific Ultraviolet Absorbance at 254 nm.

**TDN** Total Dissolved Nitrogen.

**TOC** Total Organic Carbon.

**WEOC** Water Extractable Organic Carbon.

**WEOC<sub>min</sub>** Water Extractable Organic Carbon Mineralization.

**WEOM** Water Extractable Organic Matter.

**WEON** Water Extractable Organic Nitrogen.

# Chapter 1

## Introduction

*“For most of history, few things have mattered more to human communities than their relations with soil, because soil provided most of their food and nutrients.”*

John R. McNeill and Verena Winiwarter

### 1.1 General introduction and scope

#### 1.1.1 Background

Dissolved Organic Matter (DOM) in soils is a key regulator of nutrient cycling, biological activity, and environmental health. It is considered to be the most labile fraction of soil organic matter (SOM) (Bolan et al., 2011), but several studies have contradicted this assumption inferring there is a significant recalcitrant component (Gregorich et al., 2003; Michel et al., 2006). This highly complex and heterogeneous mixture of organic compounds confers to DOM a wide range of functionalities which determine the ecological status of the soil. However, since DOM constituents cannot always be assigned to a specific chemical entity or described with a defined chemical structure, there is still a lack of knowledge about what determines its properties, availability and fate. Thus, the determination of DOM’s chemical composition and its concentration plays a crucial role in our understanding of carbon (C) and nitrogen (N) cycling in nature and, therefore, our ability to preserve or improve soil fertility, biomass production, and assess climate change risks.

Many attempts have been done to define and characterize this SOM pool: most of the studies have been done on cold and hot water extracts or other acid or salt-extractable fractions (Haynes, 2005). Such a variety of extraction protocols complicates, however, inter-study comparisons and questions the integrity of the results. Furthermore, less attention has been given to DOM in soil leachates (Ghani et al., 2010), studied foremost in forest ecosystems (e.g. Fröberg et al., 2005), and

even less is known about the changes occurring over the leaching course. There is a need, therefore, to understand the effects of soil dynamics on DOM properties and identify authentic fractions reflecting the responsiveness to agricultural practices and environmental changes.

Concentrations and fluxes of potentially mineralizable DOM are often determined by bioassay laboratory experiments using native microbial communities (see McDowell et al., 2006). Microorganisms consume and release organic fractions of C and N (Marschner and Kalbitz, 2003; Kallenbach et al., 2016), which can undergo mechanisms of adsorption and mineral protection (Sollins et al., 1996; Kalbitz et al., 2000), preserving available and potentially mineralizable DOM and preventing losses, or providing, in turn, a new valuable source for nutrient cycling. It is, therefore, commonly accepted that the amount and biodegradability of these DOM compounds play a central role in processes of mineralization and immobilization of SOM. Nevertheless, relatively little is known about microbial interactions and the reworking of DOM. In addition, the lack of a comprehensive insight into the molecular identities involved and the non-standardized nature of the soil inoculum, also hamper inter-study comparisons and lead to inconclusive answers. Therefore, it is important to find alternative approaches to predicting the DOM mineralization potential.

### 1.1.2 Research objectives

The overall objective of this PhD study was to provide an insight into the factors/mechanisms affecting DOM release and biodegradation in agricultural soils with respect to its physicochemical characteristics.

In view of the outlined research gaps, the present study sought to:

- Investigate the quantitative and qualitative changes of DOM progressively leached from a soil exposed to varying degrees of soil drying before rewetting;
- Identify functionally and biochemically distinct DOM fractions in response to land use changes;
- Determine what characteristics of DOM best describe the variability in the mineralization response;
- Identify tools and proxies of DOM biodegradability that can overcome inconsistencies and limitations brought about by sample manipulation and laboratory incubations.

### 1.1.3 Thesis structure

The present body of work comprises of six Chapters.

Chapter 1 provides a general background of the research topic, states the objectives pursued and reviews knowledge and knowledge gaps at the basis of this study.

Chapter 2 describes the fundamental concepts of the techniques employed to meet the appointed objectives and states the reasons behind their choice. Further detail on these and their practical implementation in this research work are described in the following experimental Chapters.

Chapter 3 presents the first experimental work performed with the aim of investigating the effect of the extent of soil drying prior to rewetting on the quantitative and qualitative characteristics of sequentially leached pore volumes of an arable and a grassland soil. It, ultimately, opens the debate around what aspects of DOM determine its biodegradability. The overall goal seeks to clarify how the antecedent soil water content affects the quantity and quality of DOM released when soils are rewetted, and the potential for soil carbon losses.

Chapter 4 comprises of two main Sections. The first and central part of this Chapter investigates the effect of land use and management on SOM fractions extracted with water at different extraction temperatures (cold and hot) and separated by means of anion exchange resin treatment. The objectives were to identify and characterize soluble organic matter fractions of differing reactivity, hence biodegradability. A suite of fluorescence indices was added to the UV measures in order to determine structural properties, origin and degree of decomposition. Biodegradability was assessed by laboratory incubations and these data were included in a modeling approach aiming to identify the properties that best describe the variation in the mineralization response. A first attempt towards the identification of biodegradability predictors. The outcome of this study provides new information about the qualitative aspects of DOM fractions and sub-fractions in relation to the mineralization response, and promotes the utilization of fluorescence in biodegradability studies.

The second part presents the outcome of an experiment performed ahead of that presented in the first Section. The objectives were to examine the effect of pH on the retention of DOM by anion exchange resin and to determine the optimal resin:extract ratio for the maximal DOM retention efficiency. The pH of extracts from a soil with a higher SOM content was altered within the range of pH values commonly found in agricultural soils. The retention efficacy by anion exchange resin was then assessed using, in addition, several resin:extract ratios. These data enabled the set up of the anion exchange fractionation method for the study of DOM biodegradability.

Chapter 5 focuses on the molecular properties of DOM fractions and sub-fractions and their effect on microbial activity. The first step involved a fluorescence

fingerprinting of DOM quality by means of excitation-emission matrices. The overlaying spectral components of these matrices were singled out by PARAFAC modeling and investigated, within a data mining framework, for their ability to describe DOM biodegradability and biochemical properties. The second step involved DOM profiling through a non-targeted LC–MS analysis and the classification of resolved compounds. Ultimately, the selected PARAFAC components were associated to DOM’s molecular properties and linked to the mineralization response. The results provide new insights into the qualitative aspects and mechanisms regulating biodegradability processes.

Chapter 6 conveys the knowledge gained from this study and points at the next steps to be considered in future research on this topic.

## 1.2 Literature review

### 1.2.1 DOM in natural and agricultural ecosystems

Even though it still remains an operationally defined organic matter (OM) fraction, whose concentrations are very small (0.05–0.4%; Haynes, 2005) compared to total soil organic matter (SOM), dissolved (DOM) or water extractable (WEOM) organic matter (Embacher et al., 2007) acts as a key regulator of soil carbon (C) and nitrogen (N) cycling in terrestrial ecosystems (Kalbitz et al., 2000).

DOM is defined as being organic molecules and colloids  $<0.45\ \mu\text{m}$  in diameter that are known for their ability to supply nutrients to microorganisms and plants (Bolan et al., 2011; Kaiser et al., 2015), to bind nutrients, heavy metals and organic contaminants (Zsolnay, 1996; Dudal et al., 2005; Bolan et al., 2011), and to alter microbial activity (West et al., 1992; De Nobili et al., 2001; Fanin et al., 2014). The wide range of functionalities demonstrated by DOM has been attributed to its heterogeneous chemical composition and physical properties (Thurman, 1985; Stevenson, 1994). These characteristics, however, complicate the separation of fractions with different turnover times, which remain poorly defined due to challenges in structural identification/characterization and their interaction, at the micro-scale, with a wide range of factors.

Commonly, DOM fractions are defined with respect to their stabilization, and thus extraction procedure, involving water or aqueous solutions with different ionic strength. In this context, the term WEOM specifically refers to the DOM fraction obtained by extracting SOM with water, whereas DOM most commonly reflects the *in situ* release of SOM (Zsolnay, 1996, 2003). The research described in this thesis addresses aspects of both DOM and WEOM, as both fractions have been linked to important soil processes, including many associated with C and N cycling in soils. For the purposes of this review, the term DOM is used generically to refer

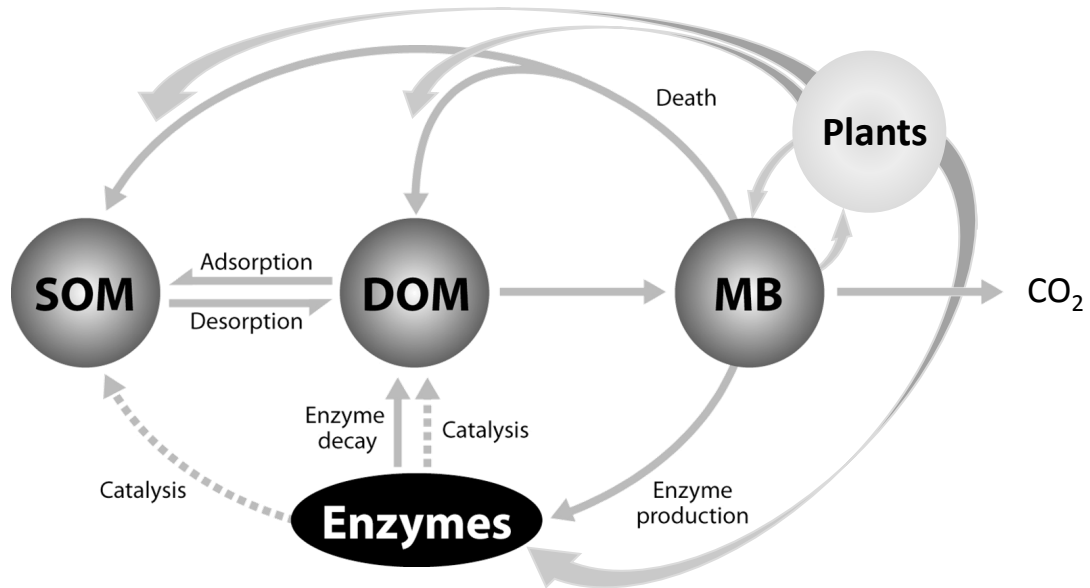
to various forms of water soluble organic matter fractions.

DOM is considered to be directly assimilable by soil microorganisms (Marschner and Kalbitz, 2003) and, therefore, its bioavailability, here defined by its hydrophilicity and molecular structure, represents a critical step in C and N biogeochemical cycling. It is generally accepted that the DOM pool consists of easily decomposable organic substances such as carbohydrates, organic acids, proteinaceous compounds, amino sugars, fatty acids and a large pool of undefined organic compounds (Poirier et al., 2005). However, Qualls and Haines (1992) and Gregorich et al. (2003) have suggested that a significant proportion of DOM comprises of recalcitrant material that is resistant to mineralization. This was true particularly for dissolved organic C (DOC) moieties. Thus, soil extractable DOM consists of both readily mineralizable and refractory OM pools. This raises the question as to what determines the biodegradability of this OM fraction, and therefore its ecological significance.

DOM is actively cycling in the soil compartment, reacting with its mineral components, being in a constant equilibrium with SOM, thus rapidly reflecting alterations in soil conditions (Zsolnay, 1996; Chantigny, 2003; Bolan et al., 2011). While much attention has been paid to DOM immobilization vs. mineralization and its losses as  $\text{CO}_2$  and  $\text{N}_2\text{O}$  in the atmosphere, less attention has been paid to the controls and properties of DOC and DON leaching from agricultural soils (Ghani et al., 2010) and their inclusion in predictive models of C and N dynamics. From this point of view, losses of DON from agricultural systems (Murphy et al., 2000) via convection, diffusion and dispersion (van Kessel et al., 2009) could have important environmental consequences, both for soil fertility and groundwater contamination, especially if the chemistry of DOM compounds is susceptible to microbial degradation. Presently, little information exists on DOC and DON losses from pasture and arable cropping soils in New Zealand (Ghani et al., 2007).

Despite its importance in nutrient mobilization and a wide range of analytical methods used to characterize DOM chemistry and functionality, there is still a knowledge gap with respect to DOM composition, biodegradability, and the factors that affect its supply. This is especially marked in field studies conducted in agricultural and pasture systems, where available results are few in number and contradictory (Chantigny, 2003; Kalbitz et al., 2003), since most of the available knowledge comes from field studies conducted in forest ecosystems (e.g. Guggenberger et al., 1994b; Fröberg et al., 2005; Jaffrain et al., 2007).

It is, therefore, clear that “DOM is not DOM” (Zsolnay, 2003): the same operationally defined DOM fractions vary in different types of environment and may perform very different ecological functions. Also, due to practical purposes or specific research questions, non-standardized methodologies of DOM extraction may lead to different conclusions. Therefore, attention must be paid to comparative and interdisciplinary studies (Zsolnay, 2003).



**Figure 1.1:** Conceptual model of the C mineralization process (modified from Curtin et al., 2012): SOM- solid-phase soil organic matter; DOM- dissolved organic matter; MB- microbial biomass.

### 1.2.2 Factors and sources of DOM release

#### Soil water content, drying and rewetting

Soil drying and rewetting cycles have been identified as being among the most important naturally occurring factors affecting DOM release. These cycles are also fundamental processes in common agricultural practices, influencing OM fate and mobility in soils. Previous studies (Kowalenko et al., 1978; Howard and Howard, 1993; Harrison-Kirk et al., 2013) suggest that these events strongly affect C and N mineralization and emissions of CO<sub>2</sub> and N<sub>2</sub>O into the atmosphere, but the effect of dry/wet cycles on SOM solubilization and biochemical composition of DOM in relation to the mineralization response is still poorly understood.

Birch (1958) observed that rewetting of a dried soil one or more times causes a flush in the CO<sub>2</sub> evolution (the “Birch effect”). Many authors (Fierer et al., 2003; Chow et al., 2006) have reported that this flush decreases with consecutive drying/wetting events. The source of the OM mineralized has been attributed to the mineralization of intracellular solutes (Halverson et al., 2000) accumulated during the stress event and released upon the rapid decrease in water potential (Kieft et al., 1987; Fierer et al., 2003; Zornoza et al., 2007), the breakdown of microbial cells (Bottner, 1985; Van Gestel et al., 1993a), or the enhanced microbial utilization



of SOM released by aggregate disruption, desorption or redistribution (Denef et al., 2001a,b; Cosentino et al., 2006). However, less is known about the changes in the physical conformation and chemical characteristics of SOM and DOM related to the drying effect.

#### *Oxygen availability and soil water effects*

Soil heterotrophic respiration is largely influenced by soil water content (Linn and Doran, 1984; Moyano et al., 2013), and so is the mineralization of DOM (Fierer and Schimel, 2002; Chow et al., 2006). In turn, the release of DOM as available substrate for soil biota is also strongly regulated by soil moisture (Zsolnay et al., 1999). Broadly speaking, soil water content is inversely related to O<sub>2</sub> availability and it is the proportion of these two parameters in the soil that determines the physical, chemical, biochemical and, therefore, physiological processes that occur. Fluctuations in soil water content will, therefore, have both direct and indirect effects on the bioavailability and biodegradability of DOM.

By affecting soil aeration, dry/wet cycles change the microbial activity and community structure (Fierer et al., 2003). This partially explains the changes in C and N mineralization rates observed following repeated drying/rewetting events. As soil pores are being emptied, water becomes progressively more disconnected (Moyano et al., 2013), thus restricting microbial activity to microscale sites at the more negative matric potential thresholds (Manzoni and Katul, 2014). As a consequence, increased soil O<sub>2</sub> availability, defined by its diffusivity through air rather than water (Cook and Knight, 2003), might promote chemical oxidation of SOM fractions encountered and determine the activity of aerobic microorganisms (i.e. nitrifying bacteria). Repeated dry/wet cycles may enhance O<sub>2</sub> penetration due to soil structural rearrangement, promoting oxidative decomposition of even more refractory material.

Rodrigo et al. (1997) inferred that the majority of ecosystem models designed to simulate C and N transformations adopt a linear response function between soil water content and C and N mineralization rates below field capacity. If the variability of soil moisture levels drives soil biotic and abiotic processes to a new equilibrium for C and N transformation rates, or the recovery of the initial state of the system is very slow (Fierer and Schimel, 2002), then the intensity of drying prior to a rewetting event may play a fundamental role in developing models of soil C and N dynamics.

#### *Microbial activity*

Several studies have observed the ability of microbial communities to rapidly adapt to changes in the soil environment, especially in soils experiencing recurring drying/rewetting cycles when compared to those from soils where moisture conditions are more constant (Fierer et al., 2003; Steenwerth et al., 2005; Schimel et al.,

2007). Many other studies have reported a flush of microbial activity after rewetting a dried soil, presumably due to the utilization of microbial debris, in addition to the OM released following the physical disruption of soil aggregates, after restoring the original conditions (Denef et al., 2001a,b; Cosentino et al., 2006). In their experiment, Wu and Brookes (2005) showed, however, that a large fraction (about 60%) of the CO<sub>2</sub>-C released derived from non-microbial biomass OM, suggesting a primary involvement of abiotic processes in DOM mineralization (Kemmitt et al., 2008).

Yet, in spite of external perturbations of the soil conditions, there might be a shift in the microbial community structure still carrying out similar functions (“redundancy principle”; Stres and Tiedje, 2006) or active in OM turnover after the release of readily available DOM compounds (“excessive pool principle”; Morris and Blackwood, 2007). In their studies, Schimel et al. (2007) suggested a selection towards Gram positive bacteria and fungi instead of Gram negative bacteria after extreme drying and fast rewetting of the soil. This is in accordance with Harris (1981), who attributed the ability of microorganisms to withstand rapid fluctuations in soil water matric potential to greater cell wall thickness. This shift in the microbial community structure could lead to a microbial group-specific organic-compound decomposition in dried-rewetted soils compared to field-moist conditions.

However, substantial proportions of microbially derived DOM have been found to increase after a drying/rewetting event, and the chemistry of these constituents is likely to change the characteristics and the ecological functions of the DOM fraction. Several authors found a strong contribution of microbially derived hydrophilic compounds (Kaiser et al., 2001), such as proteins (Miltner et al., 2009), amino sugars (Gregorich et al., 2003), contributing to the fast-cycling N-rich (Pörtl et al., 2007) DOM fraction; but, it has been argued (Kaiser et al., 2015) that the increase in the proportion of these DOM constituents could lead to an overestimation of the “assimilable” DOM pool in excessively dried samples compared to the natural soil environment.

## Land use and management

The quality and size of the DOM pool are also shaped by changes in land use and management, including vegetation cover, cultivation, grazing, fertilization and amendments (Chantigny et al., 1999; Chantigny, 2003; Embacher et al., 2008; Ohno et al., 2009). Land conversion and management practices affect primarily the SOM, and therefore DOM, in the topsoil (Haynes, 2000). Following cultivation, changes in soil structure and porosity affect the release of inter-aggregate (Golchin et al., 1994) or mineral-associated DOM (Cambardella and Elliott, 1992), thus altering both the concentration and the composition of DOM in solution. Studies conducted by Kalbitz et al. (1999) reported increases in the degree of aromaticity of

water-soluble fulvic acids in long-term arable lands as a consequence of accelerated SOM decomposition and/or little aggregate protection, in contrast to relatively little decomposed SOM in long-term intensive grasslands. It has also been demonstrated that conventional tillage practices accelerate DOM turnover and microbial decomposition (e.g. Linn and Doran, 1984), in contrast to no-tillage practices (e.g. Bayer et al., 2000; Grandy and Robertson, 2006). In addition, recent findings also suggested that increases in SOM are likely due to greater belowground OM inputs, relative to the degree of cultivation (Kong and Six, 2010; Kätterer et al., 2011; Mazzilli et al., 2015).

In mown grasslands, only a small portion of the above-ground plant biomass enters the soil system, limiting the quantity and affecting the quality of the plant litter involved in SOM decomposition processes. Also, management of grasslands has a remarkable impact on rhizodeposition, root respiration and microbial turnover (Wardle et al., 1999; Hamilton III and Frank, 2001; Klumpp et al., 2009), which are known to strongly influence mechanisms of SOM stabilization or destabilization (Cheng and Kuzyakov, 2005; Shahzad et al., 2012; Shahbaz et al., 2017).

Long-term fertilization studies have been shown to affect the quality of DOM, by examining the ratios between O-alkyl-C and aromatic C (Li et al., 2015). Variations in their proportion have been also correlated to the composition and activity of the microbial community (Michel et al., 2006). Given the strong dependency of microbial decomposition on the available substrate and the changes induced by soil management practices (e.g. changes in soil pH, redox conditions; Tian et al., 2012), increasing levels of mineral nutrients (mainly through fertilization) may reduce the production of lignin-degrading enzymes (Reid and Deschamps, 1991), which may result in an accumulation of lignins. On the contrary, depletion of nutrients (for plant growth) may stimulate the microbial utilization of the physicochemically stabilized and/or recalcitrant organic C pool (priming effect, Fontaine et al., 2011), thus promoting the release of aromatic DOM moieties as a result of microbial metabolism (Michel et al., 2006). Furthermore, under N limitation (Redmile-Gordon et al., 2015), soil bacteria produce films of negatively charged biopolymers (EPS), mainly composed by carbohydrates, proteins, glycoproteins and large amounts of extracellular DNA (Flemming et al., 2007; Dominiak et al., 2011), which: i) hold up a high level of enzyme activity (Flemming et al., 2007) capable of converting insoluble SOM into soluble forms (DOM); ii) being amphiphilic, may allow the solubilization of hydrophobic substances (by hydrophobic interaction); and iii) when released into the soil solution, may serve as a substrate for mineralization.

It is clear that the multitude of effects caused by changes in land use and management require suitable indicators of DOM properties in response to these changes, and thus suitable methods that help to characterize its different properties and biodegradability potential.

## Plants as a source of DOM

Plants have been recognized as the main contributors to DOM cycling (Kalbitz et al., 2000; Bolan et al., 2011). Plants can contribute to DOM formation directly through leachates from plant litter (Moore and Dalva, 2001; Müller et al., 2009; Fanin et al., 2014) or via decomposed plant litter/residues involved in SOM formation (Kögel-Knabner, 2002). There remain, however, uncertainties about the biodegradability of these inputs and their effect on the rate of formation of the DOM pool.

Shahbaz et al. (2017) corroborated previous studies demonstrating that (high-rate) addition of plant residues enhances macroaggregate formation (e.g. Six et al., 2004) via stimulation of microbial activity, albeit decreasing the stabilization of newly supplied aboveground C within aggregates. It has also been hypothesized (De Nobili et al., 2001) that the fresh organic matter supplied from rhizodeposition, thought to be mainly composed of low molecular weight organic acids and labile organic compounds, stimulates microbial activity and growth, and acts as a primer (Kuzyakov et al., 2000; Cheng and Kuzyakov, 2005) for SOM decomposition and DOM solubilization, providing a stimulation of nutrient cycling (Badalucco and Nannipieri, 2007). For example, priming of SOM decomposition may lead to an increase in ammonium ions and the dissolved CO<sub>2</sub> concentration in the soil solution, which may additionally enhance microbial activity.

Although the concentration of this soluble C pool is low, if compared to the OM content of the bulk soil, its importance in exerting a regulatory influence on SOM decomposition and associated N transformations has been increasingly recognized. Cheng and Kuzyakov (2005) reported accelerated rates of SOM decomposition up to 382%, along with inhibitory effects of about 50%, depending on the kind of plant-soil interactions and experimental conditions. However, the limited information on the chemical nature of the DOM released from root exudates and on the effective microbial use efficiency of the latter still does not explain the mechanistic relationship between the rhizosphere functions, the priming of SOM and processes of DOM mineralization or immobilization.

Furthermore, earlier studies conducted by Merckx et al. (1987) showed that plant nutrient supply (especially mineral N) may change the quality and quantity of root exudates, which in turn may alter the activity and the structure of the microbial population, with direct consequences on DOM cycling and composition (Chen et al., 2018). In addition, Chantigny et al. (1999) noted a positive and linear response between fertilizer mineral N concentrations and C mineralization from plant-derived materials, measured as net <sup>14</sup>CO<sub>2</sub> production. In this instance, the positive relationship was likely to be limited to the labile C pool. However, an inverse relationship has been observed with respect to decomposition of refractory C (Liljeroth et al., 1994).

### 1.2.3 The relationship between DOM mineralization and its properties: bioavailability and biodegradability

Bioavailability has been defined as the potential for soil microorganisms to interact with DOM (Marschner and Kalbitz, 2003). But, it does not imply that the microorganisms actually utilize and finally mineralize the DOM, the latter defined as biodegradability (Hodson et al., 2011). These two characteristics of DOM have often been interchangeably used with the assumption that microbial assimilation of organic compounds is directly related to their potential for mineralization. However, a distinction between bioavailable and biodegradable OM and its quantification is still a complex issue in the study of C and N fluxes.

Bioavailability is generally assessed by measures of DOM extractability, thus clearly suggesting that a comprehensive indicator of this property has not been established to date (Hodson et al., 2011). In spite of this, concentrations recovered with a defined extraction procedure are most common measures. Measures of CO<sub>2</sub> evolution, along with DOM disappearance, have been commonly accepted as indicators of DOM biodegradability and microbial activity (McDowell et al., 2006). In addition, as a result of biodegradation, (bio)synthesis of new compounds (transformation or re-synthesis) ought to be considered. These processes include both intra- and extracellular decomposition of DOM, and are both biotically and abiotically driven.

In their conceptual model for understanding the mechanisms and controls of SOM stabilization and destabilization, Sollins et al. (1996) presented accessibility, recalcitrance and interactions to be the three main characteristics determining SOM, and thus DOM, stability or biodegradability. Clearly, DOM needs to be (bio)accessible by soil microbiota in order to be bioavailable (Dungait et al., 2012). In DOM, bioaccessibility is, hence, mainly related to the physical impediment of microbes (and other soil organisms) and their ability to access DOM, thus hampering its potential bioavailability at a given time (Semple et al., 2004). Various factors may hinder microbial access to DOM, such as its protection within soil aggregates or association with mineral surfaces (Sollins et al., 1996; Kleber et al., 2007; von Lützow et al., 2007). Adsorption onto mineral surfaces will, thus, affect DOM bioavailability, but the reversibility of this reaction is controlled by exchange equilibria, along with entropy changes involving hydrophobic interactions (Kaiser and Zech, 1997). Such an impediment may also be a consequence of water disconnection in soil pores (Moyano et al., 2013) and related changes in matric potential restricting water (thus solutes) to microscale sites (Manzoni and Katul, 2014).

Chow et al. (2006) observed that the rate of CO<sub>2</sub>-C production was proportional to the DOC concentration (first order kinetics), but did not find a correlation between CO<sub>2</sub>-C and DOC in wet/dry cycles. Other studies reported a very strong relationship between DOM, inferred as WEOM, and C mineralization across a wide

range of soils and different land uses (e.g. McNally et al., 2018). Yet, Lundquist et al. (1999) stated that DOC is not a reliable indicator of C available to microorganisms, because both SOM and DOM can be substrates for CO<sub>2</sub> production, and the controlling factors for their biodegradability are directly related to substrate quality and environmental conditions (Marschner and Kalbitz, 2003).

Drying/rewetting events, for instance, have been shown to decrease DOC mineralization with time (Chow et al., 2006). If it is considered that rapid rewetting leads to an immediate release of organic cell solutes (Kaiser et al., 2015), such as amino acids, prokaryotic quaternary ammonium compounds, eukaryotic sugars, polyols, and fungal derived glycerol (Halverson et al., 2000; Fierer et al., 2003), which may be recognized as readily biodegradable, decreased mineralization may be related to reduced amounts of organic C released after each recurrent cycle. But, previous studies (Zsolnay and Steindl, 1991) attributed this to an increasing proportion of DOC which is not biodegradable. Alternatively, Warren (2014) observed a strong decrease of organic N monomers, sugars and sugar alcohols after rewetting, but he attributed their low concentration to plant or microbial uptake.

As shown by Mueller et al. (2014), bioavailability of SOC from individually incubated particle size fractions largely exceeded that inferred by incubating bulk soils; yet, the turnover of mineral-associated SOC was comparable to chemically more refractory particulate OM. Angst et al. (2017) built upon these findings emphasizing the role of plant-derived OM in mechanisms of stabilization involving aggregation and mineral surface binding. Stable isotope research conducted by Gregorich et al. (2000) and Sollins et al. (2006) also indicated a reduced potential for isotope fractionation of the heavier fractions, which were approximating steady-state C dynamics, hypothesizing a mechanism of continual microbial reprocessing of SOM (Mueller et al., 2014).

A growing body of research has reconsidered the longstanding assumptions on the turnover, and the related mechanisms, of the main DOM constituents (Kleber et al., 2011). Examples are given by lipids, commonly found in DOM isolates and typically considered relatively refractory to mineralization (Filley et al., 2008; Pisani et al., 2014), lignin, whose mineral stabilization and recalcitrance were dismissed (Kiem and Kögel-Knabner, 2003; Rumpel et al., 2004; Fuchs et al., 2011), or humic substances. For instance, the disappearance of cutin and suberin constituents has been frequently observed (Nierop et al., 2001). Evidence also suggested plant-derived lipids to have shorter turnover times than microbially derived lipids, particularly in cultivated soils (Poirer et al., 2006). Lignin degradation and pathways of its mobilization to DOM formation have also been critically examined (DiDonato et al., 2016; Klotzbücher et al., 2016).

In view of these findings, Schmidt et al. (2011) advanced the concept of persistence of such compounds in the soil sphere in response to biotic and abiotic

properties of ecosystems rather than their intrinsic molecular properties. Nevertheless, while soil mineralogy and mineral sorption were shown to have important bearings on microbial-mineral interactions (Mueller et al., 2014) and quantitative changes of DOM (i.e. bioavailability), they were attributed little importance in determining the biochemical composition of DOM (Sanderman et al., 2014). Hence, the molecular diversity of DOM along the soil profile ought to be attributed to other factors, such as microbial processing and quality of plant inputs and DOM precursors (Malik and Gleixner, 2013; Bandowe et al., 2019), but also exoenzymatic activities and abiotic processes (Sanderman et al., 2014; DiDonato et al., 2016).

Further, reductions in SOM molecular size as a result of biodegradation processes do not necessarily correspond to greater solubility: while solubility is a function of polar and ionizable functional groups resulting from hydrolysis and oxidative reactions which do lead to size reduction, low molecular weight compounds are, in fact, not equally soluble and/or reactive, suggesting the importance of chemical properties of the molecule and the conditions of the aqueous medium in determining DOM's fate.

Hydrophilic molecules with low free enthalpy and oxidation state close to zero are thought to lead to a faster microbial utilization (Kuzuyakov et al., 2015). Nevertheless, Volk et al. (1997) explained that the relationship between hydrophilic compounds, such as carbohydrates, and DOM degradation should be considered carefully, as carbohydrates can also be bound to chemically more refractory DOM compounds, such as aromatics, typical of hydrophobic DOM moieties (Guggenberger et al., 1994b). Earlier studies, yet, suggested the -OH group in position C<sub>4</sub> of the aromatic ring of lignin monomers, such as ferulic (involved in carbohydrate-lignin crosslinking) and coumaric acids, plays a larger role in the uptake and degradation of these lignin monomers with respect to e.g. cinnamic acid, which is not oxidized.

In view of the above mentioned, biodegradation of aromatics (Almendros and Dorado, 1999; Bandowe et al., 2019) and the likely microbial origin of mineral-associated DOM constituents (Kleber et al., 2011; Mueller et al., 2014) allude to the critical role of soil microorganisms in controls of DOM cycling. Notwithstanding the importance of organo-mineral interactions and aggregation in mechanisms of SOM and DOM stabilization and preservation from microbial degradation, evidence of the biochemical and structural properties in regulating DOM function is further catching on. New findings suggest the involvement of charge assisted (near-covalent) H-bonds (Gilli and Gilli, 2009; Ni and Pignatello, 2018) to explain cohesion and thermodynamic stability of soluble OM (Wells and Stretz, 2019), offering potentially new definitions of DOM formation and persistence. Hence the need to elucidate molecular properties of DOM in relation to their impact on bioavailability and biodegradability.

It is still unknown whether there is preferential microbial utilization of certain compounds and just what the characteristics of these compounds are. At the same time, the effect of drying and rewetting or changes in agricultural management regime on the formation and/or loss of DOM compounds are unknown with respect to the subsequent fluxes of mineralization and immobilization. Whether we attribute DOM biodegradability to the hydrophilic character of DOM moieties or SOM ability to oxidize and interact with polar molecules, its solubility remains a prerequisite for diffusion through microbial cell membranes (Marschner and Kalbitz, 2003) and its further biochemical pathway. Once in solution (suspension), any interaction with the mineral matrix that could hamper DOM bioavailability and influence biodegradability can, therefore, be excluded. Hence, this study will focus on the factors influencing the release and mineralization of DOM in leachates and extracts of agricultural soils with respect to its physicochemical properties.



# Chapter 2

## Materials and Methods

*“Land really is the best art.”*

Andy Warhol

Unraveling the complexity and heterogeneity of DOM represents a challenging task for soil scientists. Several complementary approaches are usually adopted to infer DOM characteristics and functionality. The main methods utilized to pursue the objectives of this thesis are described herein.

### 2.1 General chemical analyses

#### 2.1.1 Elemental analyses: total carbon and nitrogen

Determination of the total soil carbon (TC) and nitrogen (TN) contents provides an initial direct measure of the effect of land use and management on soil fertility. It also allows the relative contribution of SOM to DOM to be estimated.

Traditionally, these measurements involve either chemical (wet) or thermal oxidation methods (Carter and Gregorich, 2008). In the present study, TC and TN contents were measured by Dumas dry combustion at 1250°C in a stream of excess O<sub>2</sub> (thermal oxidation), following grinding and oven drying overnight at 60°C. These conditions allow the complete oxidation of C forms to CO<sub>2</sub> and N forms to N oxides, with the latter further undergoing complete reduction (by Cu-N catalyst) to N<sub>2</sub>. Their concentrations are then measured on a Thermal Conductivity Detector (TCD) under He carrier and expressed as % per unit mass of soil sample. Although none of the soil samples were pre-treated for inorganic C removal, the TC contents presented in this study can be regarded as soil organic C (SOC), given the negligible carbonate content of the local soils utilized.

### 2.1.2 Dissolved organic carbon

A large variety of oxidative methods has been developed to estimate the dissolved organic carbon (DOC) content, given the importance of this parameter in both soil and aquatic environments and their large scale of concentrations, often being reasonably low. One of the most common and widely accepted methods for DOC determination involves oxidation through catalytic combustion at high temperatures (680°C). Using the non-purgeable organic carbon (NPOC) method, the sample is added with acid (2 M HCl) and sparged with N<sub>2</sub> to remove any dissolved inorganic C form through the release of CO<sub>2</sub>. The sample is then passed to a combustion tube filled with a Pt catalyst heated at 680°C in O<sub>2</sub> stream, which converts all the organic forms to CO<sub>2</sub>, further detected by a non-dispersive infrared gas analyzer (NDIR). These measurements are then referred to an internal calibration curve and represent the average of three to four measurements carried out on the same sample and validated by the software. In addition, standards comprising a solution of C<sub>8</sub>H<sub>5</sub>KO<sub>4</sub> at known concentrations enable monitoring and corrections of instrumental drift for an improved accuracy of the results.

### 2.1.3 Dissolved nitrogen

Dissolved nitrogen comprises organic and inorganic nitrogen forms, the former referred to as dissolved organic nitrogen (DON) and the latter also referred to as mineral nitrogen (Min N). Since DON is generally determined as the difference between the total dissolved nitrogen (TDN) and Min N, only the methods for TDN and Min N assessment are outlined.

#### Mineral nitrogen

Inorganic (mineral) forms of N (Min N) in soils (namely nitrate– NO<sub>3</sub><sup>−</sup>, ammonium– NH<sub>4</sub><sup>+</sup>, but also nitrite– NO<sub>2</sub><sup>−</sup>– in minor amounts) are commonly recovered with a 2 M KCl extraction of bulk soils (Keeney and Nelson, 1982), but the present study mainly focused on their concentration in soil water leachates and extracts. Routine measurements of Min N follow widely adopted colorimetric methods using automated continuous flow injection analysis (FIA). This system performs a Cd reduction of NO<sub>3</sub><sup>−</sup> to NO<sub>2</sub><sup>−</sup> in an ammonium chloride matrix (Griess-Ilosvay method), which is then colorimetrically assessed following the reaction with sulfanilamide (diazotizing reagent) and *N*-(1-naphthyl)ethylenediamine dihydrochloride (coupling reagent), and a phenate reaction with NH<sub>4</sub><sup>+</sup> in the presence of an oxidizing agent and a catalyst (usually sodium hypochlorite and sodium nitroprusside, respectively), forming a blue color proportional to the concentration of NH<sub>4</sub><sup>+</sup> (indophenol blue or Berthelot reaction). Ideally, samples are analyzed shortly after

sampling and extraction, because these forms, and in particular  $\text{NH}_4^+$ , are rapidly subjected to biological transformation at optimal temperature and/or moisture conditions (Carter and Gregorich, 2008). A detailed description of these methods is given by Keeney and Nelson (1982).

### Total dissolved nitrogen

The content of available N nutrients can be determined by the alkaline persulfate oxidation method (Cabrera and Beare, 1993). Briefly, the DOM sample is added with an oxidizing solution made of potassium persulfate ( $\text{K}_2\text{S}_2\text{O}_8$ ), boric acid ( $\text{H}_3\text{BO}_4$ ) and sodium hydroxide ( $\text{NaOH}$ ), digested via autoclaving (manual method) or ultraviolet (UV)-catalysis (automated system) and colorimetrically assessed (as previously described) for  $\text{NO}_3^-$  derived from both the oxidation of DON or the solution pre-oxidation.

### 2.1.4 pH

Stoichiometric reactions in the soil solution, and thus nutrient availability, microbial activity and physicochemical properties of DOM are largely driven by the concentration of  $\text{H}^+$  ions in solution. The standard method for soil pH determination involves an extraction at a soil:water ratio of 1:2.5, followed by a potentiometric measurement of the slurry, decanted for 1 h (Devey et al., 2010), with a calibrated glass electrode, i.e. pH-meter.

### 2.1.5 Cation exchange capacity

*“Next to photosynthesis and respiration, probably no process in nature is as vital to plant and animal life as the exchange of ions between soil particles and growing plant roots.”*

Nyle C. Brady

Defined as the sum of total exchangeable cations adsorbed by soil constituents (Stevenson, 1994), the cation exchange capacity (CEC) plays a fundamental role in nutrient retention against leaching and constitutes the main source of soil mineral fertility. In the soil system, OM and clay are the major contributors to CEC, given their negative charge that attracts cations. However, the amount of negative charge and its neutralization are pH and ionic strength dependent, and thus so are cation exchange reactions. Cation displacement from the negatively charged surface depends on the charge of ions in solution (along with their hydration order) and is therefore expressed as units of positive charge per unit mass of soil ( $\text{cmol}_c \text{ kg}^{-1}$  soil).

Among a large suite of available methods, the ammonium acetate method (Thomas, 1982) is widely employed for CEC determination in low carbonate content soils and consists of a soil extraction with 1 M ammonium acetate (1:25 ratio) at pH 7 for 2 h, followed by centrifugation and assessment of cations displaced by  $\text{NH}_4^+$  (Na, K, Ca and Mg) by inductively coupled plasma - optical emission spectrometry (ICP-OES).

## 2.2 Spectroscopic techniques

*“If you want to find the secrets of the universe, think in terms of energy, frequency and vibration.”*

Nikola Tesla

Absorbance and fluorescence spectroscopies have been identified as powerful techniques being increasingly adopted in DOM characterization studies, as well as in the research chapters of this thesis. Further to their ability to provide a chemical fingerprint of DOM, they are rapid, non-destructive, and require relatively little sample, thus being particularly advantageous for timely analyses of soil leachates and extracts. The following subsections provide an overview of the basic principles of ultraviolet (UV) and fluorescence spectroscopies and their application.

### 2.2.1 Ultraviolet absorbance

The absorption of energy from electromagnetic radiations between 180 and 400 nm (near UV radiation) induces an electron transfer from  $\pi$ - or non-bonding (n) orbitals, characteristic of aromatic compounds, conjugated unsaturated species and n- electron systems. Consequently, the discrete amount of energy absorbed during the transition from ground state to excited state at a specific wavelength will result in an absorption spectra whose peak intensity enables to identify and quantify double bond bearing moieties and resonant aromatic structures of DOM, commonly referred to as DOM chromophores. In most cases, the peak intensity of a solution at a specific wavelength will reflect the sum of individual absorbances of the chromophore-containing solution components. Therefore, while robust qualitative statements can be deduced from the molecular orbitals involved, detailed structural interpretations are not plausible in this context. Furthermore, in complex mixtures such as DOM, interferences and over saturation of the signal are likely to occur; thus, it is essential in practice to adequately dilute the sample for accurate measurements.

UV measurements are a function of the cell path length; thus, recorded absorbances might not be directly comparable if not referred to a unit (1 cm) path

length. For such a purpose, the absorption coefficient ( $a_\lambda$ ) was calculated, as described by Green and Blough (1994):

$$a_\lambda = 2.303\left(\frac{A_\lambda}{l}\right) \quad (2.1)$$

where  $A$  is the absorbance at a specific wavelength  $\lambda$  (e.g. 254 nm),  $l$  is the optical path length (expressed in cm or m), and the factor 2.303 is used to convert from a logarithm in base 10 to a natural logarithm. Since the areas of maximum absorption intensity of DOM are located between 254–280 nm (Albani, 2008), a number of studies report  $a_\lambda$  as a function of  $\lambda$  at 254 nm, given the strong absorbance of aromatic compounds at that wavelength (Helms et al., 2008; Inamdar et al., 2012).

Another common parameter and the main UV indicator utilized in the research chapters of this thesis is the specific UV absorbance at 254 nm ( $SUVA_{254}$ ), defined as the UV absorbance at 254 nm normalized for the concentration of DOC and thus expressed in units of  $L\ mg^{-1}\ m^{-1}$  (Edzwald, 1993)<sup>1</sup>:

$$SUVA_{254} = \frac{A_{254}}{[DOC]} \quad (2.2)$$

Such an indicator expresses the relative contribution of absorbing moieties to the DOC in the sample and its strong correlation with aromaticity and degree of structural condensation has been validated with  $^{13}C$  NMR spectroscopy (Novak et al., 1992; Chin et al., 1994; Korshin et al., 1997; Weishaar et al., 2003; Jaffrain et al., 2007).

### 2.2.2 Fluorescence

The relaxation of molecules from their excited state to ground state may involve the emission of photons, an effect called fluorescence. Accordingly, chromophores capable of producing such effect are commonly referred to as fluorophores. Aromatic molecules (with or without heteroatoms or electron-withdrawing groups), Schiff-bases and derivatives are well known DOM fluorophores.

The processes underlying the fluorescence phenomenon are depicted in the Jablonski diagram (Fig. 2.1). Following electron excitation to a higher energy state and nuclei rearrangement towards a stable geometry (internal conversion<sup>2</sup>), the photon emission during the transition to ground state occurs at a specific wavelength

<sup>1</sup>Since the path length is commonly measured in cm, units of  $SUVA_{254}$  can also be expressed as  $L\ mg^{-1}\ cm^{-1}$ .

<sup>2</sup>This process occurs within  $10^{-12}$  s as a result of a thermal equilibration (non-radiative release of excess vibrational energy to the lowest excited vibrational level) following the instantaneous ( $10^{-15}$  s) excitation to a higher vibrational level of the reached electronic state.

independent of, and typically longer than<sup>3</sup>, the excitation wavelength (Leurgans and Ross, 1992; Lakowicz, 2006). Such an emission wavelength is characteristic of a given fluorophore; thus, it can be used for the structural characterization of DOM. However, in a molecule with multiple fluorophores, the total emission will be a function of only the fraction of the excited fluorophore that leads to photon emission; such a fraction is the fluorescence quantum yield ( $\Phi$ ), defined as:

$$\Phi = \frac{\text{number of photons emitted}}{\text{number of photons absorbed}} \quad (2.3)$$

The intensity of fluorescence emission at a given wavelength ( $I_{F(\lambda)}$ ) will, therefore, be a function of  $\Phi$ , as described by the following equation (Albani, 2008):

$$I_{F(\lambda)} = I_0(1 - e^{-2.303\epsilon cl})\Phi \quad (2.4)$$

where  $I_0$  is the intensity of the excitation beam, 2.303 is the conversion factor of the logarithm in base 10 to a natural logarithm<sup>4</sup>, and  $\epsilon$ ,  $c$  and  $l$  are, respectively, the molar extinction coefficient, the concentration of the fluorescing species and the path length defining the optical density (absorbance).

Assuming a low optical density ( $<0.05$ ), and thus a low concentration of fluorophores, such as for relatively diluted solutions, equation 2.4 can be written as:

$$I_{F(\lambda)} = 2.303I_0\epsilon(\lambda)cl\Phi \quad (2.5)$$

Simplified, it follows that  $I_{F(\lambda)}$  will be directly proportional to the number of fluorophores, and thus photons involved (Albani, 2008):

$$I_{F(\lambda)} = n \times E \quad (2.6)$$

where  $n$  is the number of photons emitted at a given wavelength and  $E$  the photon energy.

Losses occur by non-radiative decay processes (Fig. 2.1), such as fluorescence resonance energy transfer between fluorophores (FRET), inter-system crossing<sup>5</sup> and the direct interaction of the excited molecule with other molecules in solution (e.g. halide ions, oxygen, amines, referred to as quenchers), a process called quenching. Typically, quenching occurs by collision of the excited fluorophore with the quencher without molecule alteration, causing partial decrease or complete loss

<sup>3</sup>Phenomenon known as the Stokes shift.

<sup>4</sup>According to Lambert-Beer's law, the absorbance ( $A$ ) is the logarithm (base 10) of the reciprocal of the transmittance ( $T$ ). It follows that  $T$  can be expressed as  $e^{-2.303A}$ , given that 2.303 is the natural logarithm of 10.

<sup>5</sup>Conversion from singlet to triplet state with a shift towards longer emission wavelengths, a process called phosphorescence.

of the emission intensity (dynamic quenching), as described by the Stern-Volmer equation (Lakowicz, 2006):

$$\frac{F_0}{F} = 1 + k_q\tau_0[Q] = 1 + K_{SV}[Q] \quad (2.7)$$

where  $F_0$  and  $F$  are, respectively, the fluorescence intensities in the absence ( $I_{F(\lambda)}$ ) and presence of a quencher,  $k_q$  is the bimolecular quenching rate constant ( $\text{M}^{-1} \text{s}^{-1}$ ) indicating the quencher's potential to access the fluorophore by diffusion through the medium,  $\tau_0$  is the fluorophore's lifetime in the absence of a quencher (s),  $[Q]$  is the quencher's concentration (M) and  $K_{SV}$  is the Stern-Volmer constant ( $\text{M}^{-1}$ ) given by  $k_q\tau_0$ <sup>6</sup>.

An example of dynamic quenching is given by the interaction of heterocyclic aromatic compounds (e.g. quinolinium) with chloride due to electron transfer from the quencher (chloride) to the fluorophore (quinolinium). Such an effect has been considered, and phased out<sup>7</sup>, in the experiment described in Chapters 4 and 5, given the strong dependency of  $F$  and  $\Phi$  on the concentration of the quencher (i.e. chloride).

Quenchers can also form complexes with fluorophores at ground state, thus halting fluorescence (static quenching)<sup>8</sup>.

It is worth noting that DOM is itself a quencher, and according to equation 2.7, adequate dilution is a prerequisite to avoid self-quenching.

The uniqueness of excitation and emission responses for different classes of fluorophores has promoted the utilization of fluorescence spectroscopy as a reliable tool for DOM fingerprinting (Senesi et al., 1991; Korak et al., 2014). A number of fluorescence indices have been introduced to infer different aspects of DOM:

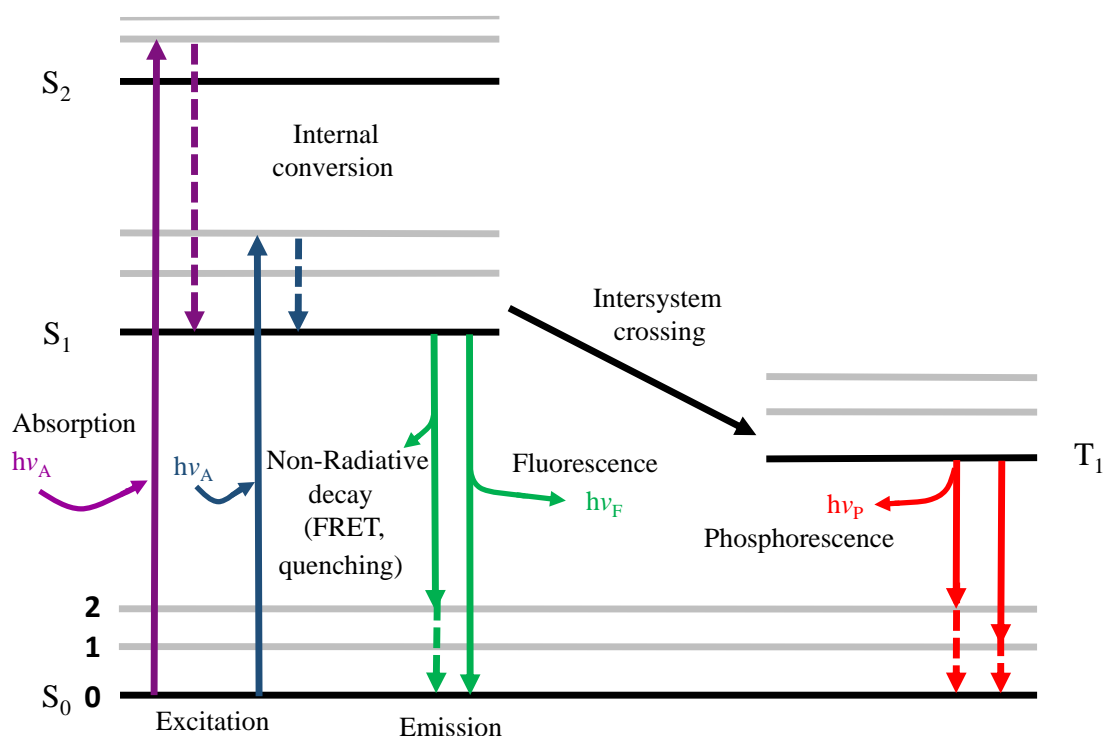
#### *DOM source*

The fluorescence index (FI) (after Cory and McKnight, 2005): originally established by McKnight et al. (2001) and defined by Cory and McKnight (2005) as the ratio between the emission wavelengths at 470 and 520 nm when excited at 370 nm, it has been largely used to distinguish microbially derived DOM from terrestrial plant and SOM-derived DOM (Jaffé et al., 2008; Inamdar et al., 2012; Fernández-Romero et al., 2016). Higher FI ( $\sim 1.9$ ) values have been associated with low fluorophores, higher aliphaticity and a lower degree of DOM structural complexity, such as DOM from microbial metabolism and cell lysis, as opposed to lower values ( $< 1.4$ – $1.5$ ) which are characterized by high fluorophores and a higher degree

<sup>6</sup> $K_{SV}$  represents the slope of the linear equation only when all fluorophores are equally accessible to the quencher.

<sup>7</sup>The maximum concentration of chloride revealed was 1.9 mM. For further details see Lakowicz (2006) and Fig. C.7.

<sup>8</sup>For static quenching,  $K_{SV}$  represents the association constant for the complex formation and does not involve a fluorescence lifetime nor affects the lifetime of the uncomplexed fluorophores.



**Figure 2.1:** The Jablonski diagram. Adapted with permission from Lakowicz (2006).

of aromaticity and structural conjugation, likely from plant decomposition and SOM reprocessing (McKnight et al., 2001; Chen et al., 2003; Cory and McKnight, 2005).

The freshness index ( $\beta:\alpha$ ) (Parlanti et al., 2000; Wilson and Xenopoulos, 2009): studies on humic substances described as  $\beta$  and  $\alpha$  the fluorophores that emit at 380–410 nm and 430–450 nm, respectively, when excited at 310–320 nm. Their ratio has been referred to the relative contribution of recently produced ( $>0.8$ ) versus highly decomposed ( $<0.6$ ) DOM, and thus is positively correlated to the FI (Wilson and Xenopoulos, 2009), given the likely biological origin of the  $\beta$  component (Parlanti et al., 2000). The application and related results concerning the FI and  $\beta:\alpha$  indices are described in Chapter 4.

#### *DOM quality*

The humification index (HIX): reflects advancing decomposition (e.g. oxidation and demethylation) and structural rearrangement (e.g. polymerization and polycondensation) of plant and animal breakdown products towards secondary synthesis of more complex, i.e. humified DOM. Such a process has been associated with a fluorescence shift to longer emission wavelengths (Senesi, 1990) due to lower resulting H/C ratios (Zsolnay et al., 1999). With respect to source related indices (FI and  $\beta:\alpha$ ), Wilson and Xenopoulos (2009) suggested the HIX to be more susceptible



to external factors, such as moisture conditions, but not land use effects.

It has been defined by Zsolnay et al. (1999) as:

$$HIX = \frac{\sum I_{435-480}}{\sum I_{300-345}} \quad (2.8)$$

at 254 nm excitation, where  $I$  represents the emission intensity at a given wavelength, and it has been further modified by Ohno (2002) to:

$$HIX = \frac{\sum I_{435-480}}{\sum I_{300-345} + \sum I_{435-480}} \quad (2.9)$$

at excitation wavelength of 254 nm, in order to avoid any potential bias deriving from the re-absorption of the emission intensity due to high DOM concentrations (the so called inner-filter effect).

The  $A_{465}$  humification index (after Milori et al., 2002): the increase in substituted aromatic structures, condensed and conjugated systems ensuing humification can be alternatively assessed as the total area under the emission spectra at 465 nm excitation. Milori et al. (2002) and Milori et al. (2006) showed an increase in resonance absorption of humified structures at this wavelength for a wide range of soils with different textures, exposed to different climatic conditions or under different land use and management. The  $A_{465}$  has, therefore, been employed as a suitable humification index for the purposes of the experiment described in Chapter 4.

The collection of multiple emission spectra over a range of excitation wavelengths projected onto a single contour map of fluorescence intensities turns into a three-dimensional excitation-emission matrix (EEM), representing the full fluorescence spectrum of the analyte. As such, EEMs have been widely utilized to study the photophysical properties of distinctive constituents of complex mixtures, such as DOM (Coble, 1996; McKnight et al., 2001; Stedmon et al., 2003; Chen et al., 2003; Baker et al., 2008; Fellman et al., 2008; Vergnoux et al., 2011; Cuss and Guéguen, 2015a). Peak locations in such matrices have been related to fluorophores of different nature and origin. Coble (1996) identified four main peaks informative of the origin and structural properties of DOM, namely (Fellman et al., 2010):

**A-peak:**  $\lambda_{ex}$  250–270 nm,  $\lambda_{em}$  380–480 nm, humic-like, terrestrially derived;

**C-peak:**  $\lambda_{ex}$  320–360 nm,  $\lambda_{em}$  420–460 nm, fulvic-like, terrestrially derived;

**M-peak:**  $\lambda_{ex}$  290–320 nm,  $\lambda_{em}$  370–430 nm, fulvic-like, microbially derived (referred to the  $\beta$  component);

**B and T-peaks:**  $\lambda_{ex}$  270–280 nm,  $\lambda_{em}$  304–312 and/or 330–368 nm, protein-like, originating from aromatic amino acids (e.g. tyrosine and tryptophan).

Yet, excitation and emission peak shifts might be overlooked due to spectral overlapping of different fluorophores. Parallel factor analysis (PARAFAC, see Section 2.5.3) has been utilized to decompose the EEM into individual spectral components (Stedmon and Bro, 2008; Murphy et al., 2013).

Although most knowledge on DOM fluorescence emerged from studies on natural waters (e.g. Hudson et al., 2007; Wells et al., 2017), a growing body of research on DOM fluorescence in soil water extracts and samples from plant origin has confirmed common fluorophores (e.g. Ohno et al., 2009; Tfaily et al., 2015; Pan et al., 2017).

## 2.3 Chromatographic techniques

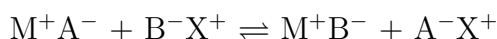
*“An expert is a person who has made all the mistakes that can be made in a very narrow field.”*

Niels Bohr

Chromatography has been defined as the separation of solution components based on their affinity for a solid phase with specific physicochemical characteristics. The usefulness of fractionation techniques in DOM studies allows the separation of DOM into groups of chemical components based on their physicochemical properties. Column and batch chromatographic experiments were performed using the techniques here described.

### 2.3.1 Anion exchange chromatography

Anion exchange chromatography builds upon electrostatic interactions occurring between the positively charged solid phase and anions in solution, according to the following equation:

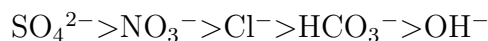


This approach turns out particularly interesting for the study of DOM, due to the abundance of functional groups negatively charged at the common range of soil pH.

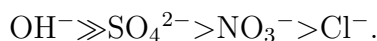
Common anion exchangers are weak-base anion resins, containing tertiary amino groups, or strong-base anion resins, with type I ( $-N^+(CH_3)_3$ ) or type II ( $-N^+(CH_3)_2(C_2H_4OH)$ ) quaternary ammonium functional groups. The advantage of strong-base anion exchange resins is that they are protonated over the whole pH range ( $pK_a > 13$ ). Further, type I resins are more stable, whereas type II resins are more hydrophobic (ethanol groups) and have a greater capacity and regeneration efficiency (Wachinski, 2016).

These positively charged functional groups are attached to a polymeric matrix (solid phase), commonly polystyrene or polyacrylic, which is either macroporous (highly porous) or gel (no pores). Polystyrene matrices are more hydrophobic, whereas the polyacrylic retain more water (Bolto et al., 2002; Boyer and Singer, 2008). Bolto et al. (2002) and Kanazawa et al. (2004) also showed that polystyrene resins were more selective than acrylic, likely involving size exclusion (Boyer et al., 2008). In addition, a macroporous resin network enables a rapid inclusion and diffusion of larger DOM molecules and solvents (Malcolm and MacCarthy, 1992). It was therefore concluded that polyacrylic, macroporous, strong-base anion exchange resins would be most effective for DOC removal (Bolto et al., 2002).

The charge of strong-base anion exchangers is compensated by counter-ions, most commonly chloride. Thus, the anion exchange occurs by displacement of chloride anions by DOM, a mechanism depending of the ionic strength of the solution and relative to DOM surface charges and dielectric constant. The selectivity of strong-base anion exchangers for most common inorganic anions has been defined to vary from (Wilson et al., 2000):



in contrast to weak-base anion exchangers, displaying an inverse selectivity:



This aspect and the relative concentration of species in solution need to be considered for the competition for sorption sites.

Anion exchange is commonly performed in batch or column experiments. In the first method, the resin is added to the soil solution and removed after the exchange has occurred, whilst in the second method, the soil solution is passed through the resin packed into a column. The latter is generally preferred in industrial applications (water treatment) due to the ease of regeneration and, thus, a larger number of reuses. For the experimental purposes of this thesis, anion exchange was conducted in batches.

Most common and effective applications of anion exchange involve waste water and landfill leachate treatments (Afcharian et al., 1997; Kanazawa et al., 2001; Bolto et al., 2002; Boyer and Singer, 2008; Bashir et al., 2010). Nevertheless, in regard to DOM and WEOM studies, the disadvantage of using strong-base resins is their tendency to irreversibly adsorb humic substances (Wachinski, 2016). For this reason, in the studies conducted in Chapters 4 and 5, whole extracts and resin treated extracts were principally discussed.

### 2.3.2 Liquid chromatography

Liquid chromatography (LC) represents the principal method for the separation of complex mixtures relative to their affinity to liquid (mobile) and solid (stationary)

phases. Due to the heterogeneous nature of DOM, two types of LC have been adopted in the study conducted in Chapter 5 and their characteristics are described herein.

### Reversed-phase liquid chromatography

One of the most commonly adopted chromatographic techniques for small organic molecules, such as peptides and small proteins, is reversed-phase (RP)-LC. Reversed-phase because the stationary phase consists of a hydrophobic (non-polar) matrix, commonly silica-based (silanol groups) and covalently bound to alkyl chains of different lengths (e.g. n-octadecyl, referred to as C18), whereas the mobile phase is hydrophilic (polar), which are inverse characteristics of the so called normal-phase (NP) chromatography. The separation occurs such that the hydrophobic moieties bind to the stationary phase in order to increase their system entropy (Tanford, 1980), thus the solvated polar moieties are eluted first. Subsequently, the polarity of the mobile phase is gradually decreased by increasing the percentage of organic solvent (e.g. acetonitrile), weakening the entropy of the hydrophobic interaction and allowing the elution of the molecules retained. The gradient applied, along with the characteristics of the stationary phase and column dimensions, are key factors influencing the chromatography, and thus the number of resolved compounds. A longer chain stationary phase usually suits better small and/or less hydrophobic molecules in order to increase their retention (Boone and Adamec, 2016); hence its frequent use in DOM studies (e.g. Sleighter and Hatcher, 2008). Importantly, C18 phases with endcapped<sup>9</sup> silanol groups are preferred in order to avoid undesired polar interactions. Column and further chromatography details pertaining the current study are described in Chapter 5.

### Hydrophilic interaction liquid chromatography

Hydrophilic interaction liquid chromatography (HILIC) is an emerging LC technique that combines the chromatographic characteristics of both RP and NP chromatographies and allows the analysis of charged molecules, generally performed by ion exchange (Buszewski and Noga, 2012). HILIC uses a hydrophilic (polar) stationary phase and a combination of an organic solvent and aqueous solvent as a mobile phase, suitable for electrospray ionization and, thus, adequate for coupling with mass spectrometry. It, therefore, preferentially retains polar, hydrophilic molecules. A gradient is applied, such that the increasing proportion of the polar (aqueous) eluant reduces the retention of hydrophilic species, which are eluted in order of increasing polarity. Practically, a mechanism resembling the NP chromatography,

---

<sup>9</sup>End-capping refers to a procedure that replaces unreacted silanol groups of the stationary phase, which may cause secondary polar interactions with the solution.

but with the advantage of retaining also small, uncharged (zwitterionic) and amphiphilic molecules due to their partitioning between the aqueous semi-immobilized layer formed on the stationary phase and the mobile phase, involving both dipole-dipole and H-bond interactions (Buszewski and Noga, 2012). These features make the HILIC versatile and particularly suitable for small polar compounds in complex mixtures (Hemström and Irgum, 2006). Common applications include analyses of peptides, nucleotides, carbohydrates, alkaloids and generally small ionizable metabolites (Buszewski and Noga, 2012). Clearly, a promising method of choice for DOM studies. Polar stationary phases vary from common unmodified silica to anionic (e.g. polyethyleneimine), cationic (e.g. carboxylate) or zwitterionic (e.g. sulfoalkylbetaine) phases, while mobile phases typically consist of aprotic solvents (commonly acetonitrile) in small parts of water. Along with RP-LC, HILIC was used in the study presented in Chapter 5.

## 2.4 Mass spectrometry

*“Good decisions can’t be made in the absence of good information...”*

Anon

To date, the chemical and structural identification of small molecular identities from complex mixtures, such as DOM, is still at its early stage. Yet, mass spectrometry (MS) performs as the best choice of technique for such a purpose, thanks to its ability to ionize DOM, its high resolving power and ease of coupling with a liquid-chromatography (LC) interface (Perry et al., 2008), phasing out sample pretreatment, and thus limiting artifacts due to pH adjustments or DOM pre-concentration. One of the best performers, the Orbitrap<sup>™</sup> mass spectrometer, played a pivotal role in the final part of this thesis (Part II). The following subsections delineate the main features of this technology and the following process of empirical molecular formulae assignment and interpretation.

### 2.4.1 High-resolution accurate-mass (HRAM): Orbitrap

High spectral resolution and mass accuracy (HRAM) in MS have been promoted by the Orbitrap<sup>™</sup> technology (Thermo Fisher Scientific). The concept of orbital trapping of ions in an electrostatic field has been introduced by Kingdon (1923) and the so called Kingdon trap. Ions enter an electrostatic orbital trap (namely Orbitrap) consisting of a spindle-like inner electrode and a differential pair of outer electrodes generating a radial logarithmic potential between the electrodes and a harmonic axial potential inducing the ions to orbit around and exhibit harmonic oscillations along the central electrode (Perry et al., 2008). The image current

induced by the axial ion oscillation is detected by the pair of outer electrodes and passed on to an amplifier. Finally, the detected signal is Fourier transformed into the ion’s mass spectrum according to its mass-to-charge ( $m/z$ ) ratio<sup>10</sup>.

At any given  $m/z$  ratio ( $m$ ), the mass resolution can be determined from the full width of a spectral peak at half-maximum peak height ( $\Delta m_{50\%}$ ), defined as FWHM by the following equation (Perry et al., 2008):

$$\text{Resolution (FWHM)} = \frac{m}{\Delta m_{50\%}} \quad (2.10)$$

The resolving power will finally affect the accuracy of mass determination, estimated as:

$$\text{Mass error (ppm)} = \frac{m_{\text{measured}} - m_{\text{true}}}{m_{\text{true}}} \times 10^6 \quad (2.11)$$

and the signal-to-noise (S/N) ratio, generally determined as the ratio of the signal peak height from the baseline to the standard deviation of the baseline.

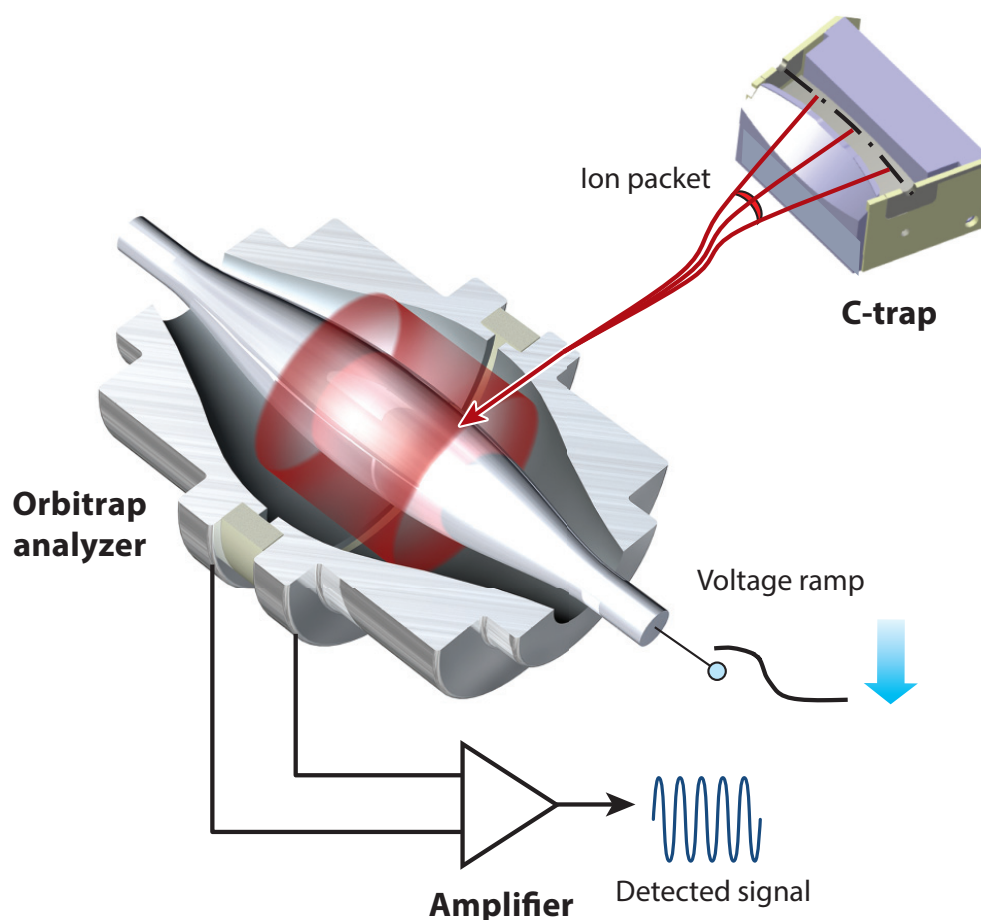
A key feature of the Orbitrap mass analyzer is an external curvilinear ion trap (C-trap) that enables ion accumulation prior to injection in the analyzer (Figure 2.2). This feature permits a pulsed transmission of ion batches, orthogonally to the mass analyzer, continuously released from an ion source, such as electrospray ionization (ESI) or atmospheric pressure photoionization (APPI); alternatively, it transmits ions axially to a collision cell, where ions are fragmented before being returned to the C-trap and ejected to the mass analyzer. However, the amount of charge filling the C-trap needs to be controlled by setting an injection time or number of pulses (known as  $IT_{\text{target}}$ ) in order to avoid space charge effects (Eliuk and Makarov, 2015).

Ionization plays a fundamental role in converting molecules into charge states and conferring kinetic energy to the formed ions towards the MS system; however, it demands careful optimization of the ionization source parameters, due to the selectivity of ESI towards polar species and effectiveness of APPI for non-polar and slightly polar species.

The fragmentation process, involving the dissociation of the ion from its neutral loss (Pluskal et al., 2012), is a great advantage for the identification of structural differences between ions (Desiderio, 1994), in combination to an additional quadrupole mass filter ahead of the C-trap<sup>11</sup> (analytical setup referred to as tandem mass spectrometry, MS/MS) enabling the pre-selection of precursor masses from the mixture of ions formed in the ion source (e.g. ESI). The high sensitivity of such a configuration (resolutions  $>150\,000$  FWHM and mass accuracies  $<1$  ppm) facilitated the

<sup>10</sup>The electrostatic force applied is a function of mass and charge according to the Newton’s second law and Lorentz force law, as is the frequency of axial oscillation.

<sup>11</sup>The the latest development of the Q Exactive model adopted for the current study.



**Figure 2.2:** Schematic representation of the Orbitrap mass analyzer and a cross section of the C-trap. The example shows the ion packets entering the mass analyzer, where the voltage applied stabilizes them into a rotational motion around the central electrode, inducing image current further detected by the amplifier. Reprinted with permission from: Eliuk and Makarov (2015), Thermo Fisher Scientific.

non-target analysis of complex mixtures, given the link established between the precursor and the product ions, thus allowing a post-analysis selection of meaningful mass peaks from the full scan data for further interpretation. Furthermore, coupling with ultrahigh pressure (UHP)–LC has significantly improved S/N ratios, advancing the analysis of low molecular weight molecules (<500 Da), and thus, NOM and DOM characterization studies (e.g. Cortés-Francisco and Caixach, 2013).

### 2.4.2 Detection of unknown compounds

In the current state of art of spectral analysis of complex mixtures, heuristic rules applied to chemometric techniques represent the basis of compound identification. The complexity of molecular formula assignment lies in the fact that several molecular formulae can be determined from one molecular mass, despite high resolution and mass accuracy. Seven golden rules established on heuristic and chemical criteria for molecular formulae prediction and validation have been detailed by Kind and Fiehn (2007). These rules were accounted for in the workflow set up (details in Appendix C) and the post-analysis data processing (see Subsection 5.2.9) at the basis of the results presented in Chapter 5.

In the first instance, the selection of relevant chemical elements and the restriction of their count for a molecule mass range reduces false predictions and computational time. Soil organic matter is mainly composed of C, H, N, O, with minor contributions of P and S, as well as other microelements. Since the number of possible combinations between these atoms varies with molecular weight (Koch et al., 2007), atom, atom count and maximum precursor mass have to be restricted. However, *a priori* exclusion of elements may cause substantial errors (Koch et al., 2007); thus, further rules apply.

Following ESI, radicals may occur as fragments of a certain molecule, that may not exist in its natural state. Thus, resolved species need to be neutralized for their adducts (e.g. mass + 1.007825 Da for a negative ion, abbr. M+1) in order to be validated by meeting the ‘octet rule’ (Lewis rule) and the Senior’s theorem conditions (Senior’s rule), the latter requiring:

- i. the  $\sum$  of valences or atoms having odd valences to be even;
- ii. the  $\sum$  of valences to be  $\geq$  twice the maximum valence;
- iii. the  $\sum$  of valences to be  $\geq$  twice the number of atoms -1,

and allowing for the maximum valence state of each element and different valence states within the molecule (Kind and Fiehn, 2007).

Evaluation of isotopic abundance patterns is a fundamental criteria for formulae computation (Kind and Fiehn, 2006), especially when attempting to identify larger and structurally more complex molecules, likely to be part of DOM. That is because different isotopes of constituent atoms (isotopologues) may coexist within the monoisotopic ion (e.g. peptides). The search generally scores mass spectral peaks for the monoisotopic ion with more than one isotopologue, including only signals with an isotopic pattern for molecular formulae prediction (Verkh et al., 2018).

Element molar ratios, such as the H/C ratio, further restrict the likelihood of a possible formula. Commonly, H/C ratios for NOM and SOM lie between 2 (e.g. long chain alkanes) and 0.5 (e.g. polycyclic aromatic hydrocarbons), but values  $4 \geq$



$H/C \geq 2$  and  $0.5 \geq H/C \geq 0.125$  are also found (Kind and Fiehn, 2007). Similarly, O/C ratios are low (e.g. 0.2–0.6) for highly aromatic molecules and increase (up to 1.2, but increases up to 2 are not uncommon) in highly saturated molecules. Kind and Fiehn (2007) reported a 13-fold reduction of computed molecular formulae by applying element ratio constraints. For particularly large organic macromolecules, predictions are refined evaluating the combination of all element ratios within the molecule (condition defined as “multiple element count” by Kind and Fiehn, 2007). It is noteworthy, however, that a number of known naturally occurring metabolites (e.g. ATP) do not meet the restrictions imposed by this criteria. In light of this and given the range of molecular weights observed in the study conducted in Chapter 5 (mainly <500 Da), this criteria was considered, yet omitted from the filtering process.

Ultimately, spectral libraries of known natural and synthetic compounds are commonly consulted as part of the workflow performed by softwares designed *ad hoc* for compound profiling (Halket et al., 2005; Kind and Fiehn, 2006). Yet, relatively accurate spectral and formulae matches do not warrant structure definition and compound annotation of unknown compounds due to a large variety of possible isomers, many of them even not available up to date (Kind and Fiehn, 2007).

### 2.4.3 van Krevelen diagrams and structural determinants

The van Krevelen diagrams provide a simple way of visualizing and representing the chemical composition of complex chemical mixtures by plotting the O/C (oxidation) versus the H/C (unsaturation) ratio for every compound identified in the sample. Such a representation broadly defines the biochemical fingerprint of DOM and enables the identification of major changes in the molecular composition occurring as a result of physicochemical and biological processes. Upon the development of HRAM techniques and convenient chromatographic separation, biochemical fingerprinting can be extended to single fractions, thus more accurately conveying the chemical information pertaining the diagenesis of the organic mixture.

Additional value of the van Krevelen diagrams is given by indices such as the double bond equivalent (DBE), indicating the number of double bonds per molecular unit and calculated as (Koch and Dittmar, 2006; Murray et al., 2013):

$$DBE = 1 + \frac{\sum_i^{i_{max}} N_i (V_i - 2)}{2} \quad (2.12)$$

where  $N_i$  and  $V_i$  are the number of atoms for each element  $i$  and its valence, respectively. Thus, for a fully saturated molecule,  $DBE = 0$ , while increases in C, N and P and a decrease in H atoms will increase DBE. Given that DOM is largely constituted by carboxyl-bearing moieties, C=O bonds are accounted for and unresolved

by DBE. Normalizing DBE by the number of C atoms in the molecule (DBE/C) renders the aromatic nature of the molecule (Hockaday et al., 2006). Yet, the variety of heteroatoms likely comprising the molecule biases the relationship between DBE/C and C=C bonds. A conceptual alternative of the DBE has been proposed by adjusting the DBE to the relative oxygen content (DBE-O), in order to provide an indication of the relative contribution of oxygenation to the degree of unsaturation of the molecule and to discriminate solely aromatic and condensed aromatic structures (Koch et al., 2005; Tfaily et al., 2013). Furthermore, an aromaticity index (AI) proposed by Koch and Dittmar (2006)<sup>12</sup> accounts for the heteroatom contribution to double bonds and can be regarded as a measure of conjugated and aromatic structures.

## 2.5 Statistical analyses and chemometric techniques

*“Any fool can know. The point is to understand.”*

Albert Einstein

### 2.5.1 Analysis of variance

The analysis of variance (ANOVA) aims at characterizing any difference in the mean between two or more groups brought about by an independent variable (categorical factor) and any external factor other than the independent variable (residual variable). Accounting for the residual variable means accounting for any error in the estimation of the main effect caused by inherited differences *within* each group. For each group, this error is expressed as the sum of squares ( $SS_e$ ) of the differences between each observed value ( $x_{ij}$ ) and the mean of its corresponding group ( $\mu_j$ ):

$$SS_e = \sum_{i=1}^n (x_{ij} - \mu_j)^2 \quad (2.13)$$

Thus, the overall effect of the residual variable is simply the sum of these group sum of squares ( $SS_E$ ).

The same concept is applied to estimate the effect of the independent variable (A), thus the differences occurring *between* the groups. In this case, the overall mean ( $\mu$ ) is subtracted from the mean of each individual group ( $\mu_j$ ) and the square of the difference is weighted by the number of groups ( $n_j$ ) and summed up ( $SS_\alpha$ ).

---

<sup>12</sup>The calculation of AI and  $AI_{mod}$  had been recently corrected by Koch and Dittmar (2016).

The sum of all sums defines the main effect:

$$SS_A = \sum_{i=1}^n n_j (\mu_j - \mu)^2 \quad (2.14)$$

The overall variation induced by the independent and residual variables is, hence, inferred by the variance ( $\sigma^2$ ), or total sum of squares over the degrees of freedom (independent measurements) for the whole data set ( $n-1$ ):

$$\sigma^2 = \frac{\sum_{i=1}^n f_i (x_{ij} - \mu)^2}{n - 1} \quad (2.15)$$

where  $f_i$  is defined as the probability ( $p_i$ ) of a number of observed values ( $n$ ) to occur multiple times.

The statistical significance of the main effect with respect to the residual effect within a defined range of probability is given by the ratio between the  $SS_A$  and  $SS_E$ , each normalized by the degrees of freedom (independent measurements) for the number of groups ( $k$ ) and the total number of observations ( $n$ ), respectively:

$$F = \frac{SS_A/(k - 1)}{SS_E/(n - k)} \quad (2.16)$$

where  $F$  denotes a test value to be compared with a tabulated  $F$  value, established by Ronald Fisher, critical for a given level of significance ( $\alpha$ ), and  $(k-1)$  and  $(n-k)$  represent the degrees of freedom for  $k$  and  $n$ , respectively. It is intuitive to conclude that, the larger the value of  $F$  for a given significance level, the larger the confidence to reject the hypothesis that there is no statistically significant difference between the group means (null hypothesis). Further to the  $F$ -statistics, ANOVA applies several other tests (e.g.  $t$ -test) in order to identify what pair of groups cause a statistically significant difference (defined by a  $P$ -value).

The close relationship between ANOVA and regression analysis (Iversen, 2011) is explained by the fact that ANOVA performs as a linear regression when all the independent variables are discrete (Shalizi, 2017). In addition, when such variables derive from *ad hoc* designed experiments, it is possible to infer the causal effect on the dependent variable. The model, described as *general linear model*, is defined by:

$$y = y_0 + ax + \epsilon \quad (2.17)$$

where the intercept of such regression ( $y_0$ ) will correspond to the mean of a reference group, whereas the slope of the regression ( $a$ ) will relate to the difference between the mean of the reference group and another group, and  $\epsilon$  is the associated residual error. In this context, the coefficient of determination, known as  $R^2$ , determines how

much of the total variation in the dependent variable is explained by the variation in the independent variable:

$$R^2 = 1 - \frac{\sum (y_{actual} - y_{predicted})^2}{\sum (y_{actual} - y_{mean})^2} \quad (2.18)$$

In Chapter 4, ANOVA has been applied to investigate the effect of land use on soil's general properties, as well as the amount of WEOM per unit of soil mass. Tables 4.1 and 4.2 provide a summary of the ANOVA output, with group means and computed standard errors of the difference of means and least significant differences among groups.

### 2.5.2 Restricted maximum likelihood and linear mixed models

Multilevel and hierarchical data structure that raise into clusters of related data and collinearity, promote the use of linear mixed models (LMMs) in place of simple linear models. Examples of such data structures are presented in Chapters 3 and 4. The characteristic and advantageous feature of LMMs is brought about by the integration of both fixed and random effects<sup>13</sup> into one unique model (thus mixed), avoiding multiple simple regressions. This not only allows the use of the whole data set (maximizes the sample size), but most importantly, accounts for the variance coming from the random effect(s). Furthermore, they are suitable for a large set of parameters, unbalanced experimental designs and missing data. The model can be summarized as (Laird and Ware, 1982):

$$y_i = X_i\beta + Z_ib_i + \epsilon_i \quad (2.19)$$

where  $y$  is an  $N$ -vector of  $i$  responses,  $X_i$  is a known  $n_i \times p$  matrix for the fixed effects,  $\beta$  is a  $p$ -vector of unknown regression coefficients for the fixed effects,  $Z_i$  is a known  $n_i \times q$  matrix for the random effects,  $b_i$  is an independent  $q$ -vector of random effects, and  $\epsilon$  is an independent and normally distributed  $N_i$ -vector of residuals (with zero mean and covariance matrix  $R$ ). The parameters  $\beta$  and  $b_i$  are, thus, subjected to estimation.

Estimating model parameters is a key step in statistical modeling. The maximum likelihood (ML) is the most common estimator of the variance component in  $\beta$  and  $b_i$ . The likelihood estimates the probability to observe a set of data as a function of a parameter (Shalizi, 2017), provided that the data are fixed. Thus,

---

<sup>13</sup>Fixed effects can be broadly described as independent variables (factors) expected to affect the response variable, while random effects as grouping factors causing collinearity among data within the same sample group.

maximizing the likelihood of the model's parameters will maximize the probability of finding the least adequate model that best fits a set of given data. In an ANOVA approach, this would be equal to minimizing the sum of squared error. However, to find ML estimates in LMMs, likelihoods over all possible values of the random effects must be integrated. As such, the restricted maximum likelihood (REML) will dissociate the estimates for the variance components from the estimates for the fixed effects and enable to assess the significance of any random effect without being affected by the fixed effects.

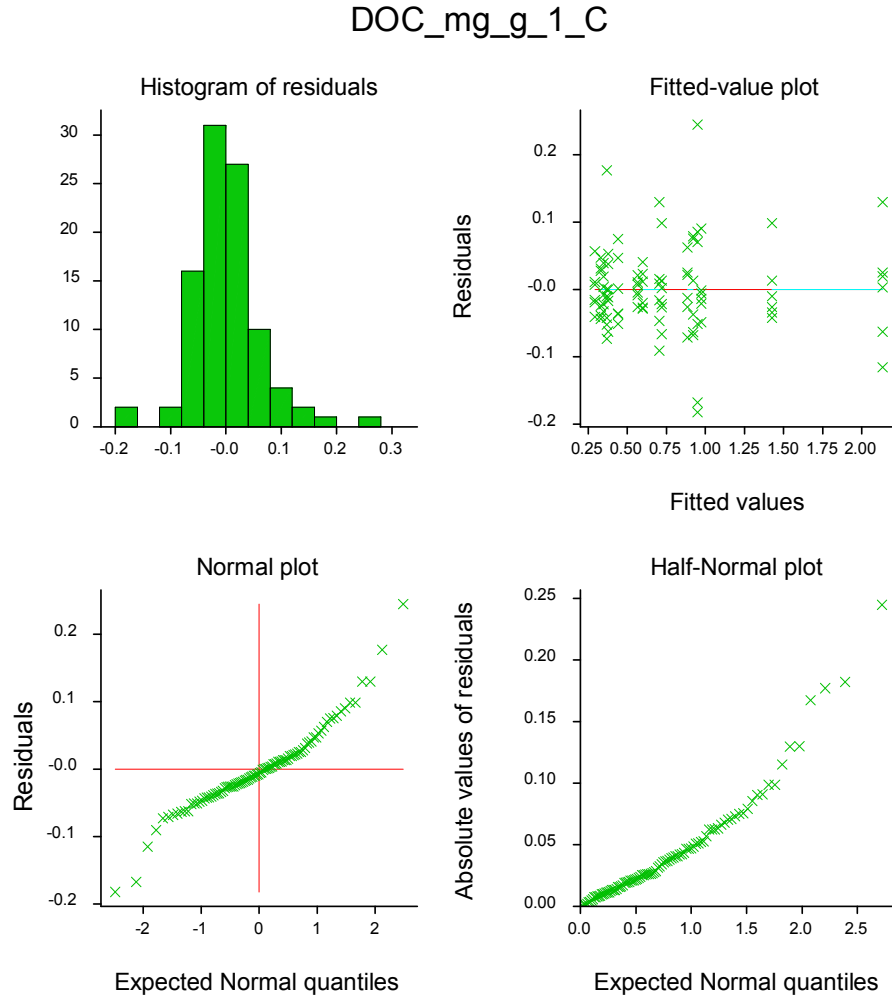
### Example: The drying and rewetting leaching experiment<sup>14</sup>

Here an example of a LMM output fitted with REML. The treatment (moisture content) and leaching number (pore volume) effects on the variable (e.g. DOC) was described through a series of fixed and nested terms and their interaction. The code was written in Genstat v.17<sup>15</sup>.

```
poin data; !p(DOC_mg_g_1_C)
calc nv=nvalues(data)
prin nv
getat[att=id]data[];id[1...nv]
prin id[i][1]
getat[att=lab]Soil;s
point id2;!p(#s[])
prin id2[]
for [index=i;ntimes=nv]v=data[]
for [index=j;ntimes=2]
restr v,Treatment,Treatment_1,Treatment_2,Pore_volume,Rep;Soil.eq.j
vcomp [fix=(Treatment_1/Treatment_2/Treatment)\
+(Treatment*Pore_volume);fact=9];con=pos
reml [prin=#,wald,dev;maxcycle=200]v
vcomp [fix=Treatment*Pore_volume;fact=9];con=pos
reml [prin=#,wald;maxcycle=200]v
vgra [pse=lsd]x=Treatment;groups=Pore_volume;title=id[i][1];ytitle=id2[j]
vlsd [print=lsd]
vplot
endf
endf
```

<sup>14</sup>Further details and results of this experiment are reported in Chapter 3.

<sup>15</sup>Genstat is a statistical software package originally developed at Rothamsted Research for data analysis in the field of agriculture and extended to physical, mathematical and natural sciences.



**Figure 2.3:** An example of residuals' plots from the LMM fitted with REML.

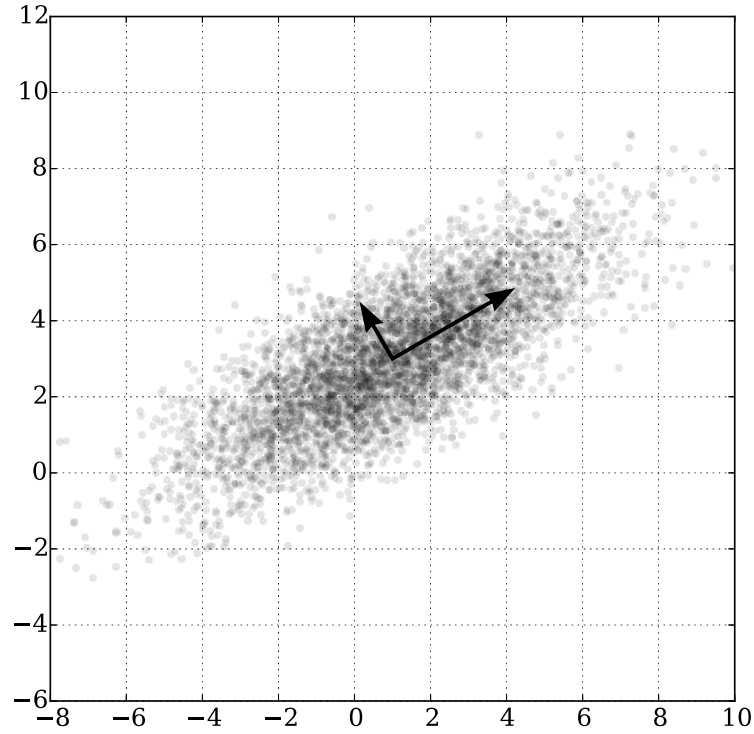
### 2.5.3 Multivariate analyses

#### Principal component analysis

Dealing with high-dimensional data, while preserving information integrity, has become feasible with another technique of the multivariate analysis family, known as the Principal component analysis (PCA).

PCA reduces the number of dimensions in which the data lie and summarizes them as a set of linearly uncorrelated variables, called principal components (Webster, 2001; Shalizi, 2017).

*Principal components* are derived so that the orthogonal projection of the original data points onto a new axis maximizes the variance of the projected values.



**Figure 2.4:** Eigenvectors of a multivariate Gaussian distribution. Reprinted from: Wikimedia Commons, <https://commons.wikimedia.org/wiki/File:GaussianScatterPCA.svg>.

The projected values corresponding to the new direction are the *principal component scores*, and the axis representing the largest variance, the first principal component. The cosine of the angle between the original and the new axis denotes the *eigenvector* of the first component, and the variances of the original points along the new axis, its *eigenvalues*. There are  $q$  principal components; each successive principal component is a normalized linear combination of the original variables ( $p$ ) that maximizes the remaining variance for that subspace and is orthogonal to the previous one (thus uncorrelated). Therefore,  $q$  generally represents the total number of original variables. The *loadings* ( $e_i$ ) are the coefficients of the linear combinations of the continuous variables. When represented within a circle with unit radius, their proximity to the circumference shows their strength in a given dimension (Webster, 2001). Summarized, the variance of the ( $i^{th}$ ) principal component ( $Y_i$ ) will be:

$$\text{var}(Y_i) = \sum_{k=1}^p \sum_{l=1}^p e_{ik} e_{il} \sigma_{kl} = \mathbf{e}_i' \Sigma \mathbf{e}_i \quad (2.20)$$

and the covariance allowing for uncorrelated components:

$$\text{cov}(Y_{i-1}, Y_i) = \sum_{k=1}^p \sum_{l=1}^p e_{i-1,k} e_{il} \sigma_{kl} = \mathbf{e}_{i-1}' \Sigma \mathbf{e}_i = 0 \quad (2.21)$$

*Pro et contra:* PCA performs as a robust descriptive, exploratory technique that does not require model assumptions (Jolliffe, 2011). Inferring the principal components reduces the number of meaningful predictors, which makes it, therefore, suitable for modeling purposes. Further, while keeping the predictors independent, it avoids multicollinearity problems, yet considering only linear relationships.

### Example: The “Millenium tillage trial” (MTT) data set<sup>16</sup>

This data set consisted of 216 observations of 45 numerical variables. The observations were grouped into four categorical factors: “Soil”, “Extraction”, “Treatment” and “Time”, corresponding to land use, extraction method, anion exchange resin treatment and pre- or post-incubation measurement, respectively. PCA was then conducted on two main data subsets: (i) variables including pre- and post-incubation measurements ( $n = 19$ ) and (ii) variables related to the pre-incubation measurements and WEOM mineralization (expressed as a cumulative measure over a time period or as the difference between the pre- and post-incubation times) ( $n = 22$ ).

In this example, PCA was performed in R<sup>17</sup> using the function `prcomp`<sup>18</sup>.

```
> PCA_time_0_finalset <- prcomp(pca_rep_time_0_finalset[, c(6:27)],
                                center = T, scale. = T)

summary(PCA_time_0_finalset)
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	
Standard deviation	3.5286	2.4825	1.31270	0.69150	0.64866	...
Proportion of Variance	0.5659	0.2801	0.07833	0.02174	0.01913	...
Cumulative Proportion	0.5659	0.8461	0.92440	0.94613	0.96526	...

<sup>16</sup>The PCA results of this experiment are reported in Chapter 5 and related Appendix.

<sup>17</sup>R is an open source software environment for statistical computing and graphics (R Development Core Team, 2018) implemented (from the S programming language) as a function-based language that compiles and runs a large variety of extendible libraries of packages.

<sup>18</sup>In R, PCA can be performed with two functions, which differ in the mathematical approach behind (Shalizi, 2017): `princomp` analyzes the *eigen-decomposition* of the variance-covariance matrix (squared matrix); `prcomp` uses the *singular value decomposition* (SVD), which reduces a rectangular matrix into its constituent parts, two orthogonal matrices and one diagonal matrix (Salkind, 2007), and is generally more robust for multivariate analyses.



Prior to actual PCA, mean and variance were “centered” and “scaled”, i.e. variable’s average was zeroed out (`center()`) and each variable was normalized to unit variance (`scale()`) to allow comparisons. This is a fundamental step when dealing with different units of measurements within the data set and when the same variable can be expressed with different units (Jolliffe, 2011).

Components were selected setting a threshold of accounted percentage of total variation equal to 80% (Jolliffe, 2011). Thus, axes were rotated only within the selected  $q$ -dimensional subspaces for exploratory and interpretation purposes. In the current example, two principal components were selected and their loadings evaluated according to Fig. C.20 in Appendix C.

### Parallel factor analysis (PARAFAC): a principle of parsimony<sup>19</sup>

Decomposition of multi-way data, i.e. sets of two dimensional matrices arranged in multiple dimensions, into unique components capable of describing the full data array represents a suitable method for the interpretation of trilinear structures<sup>20</sup> (Bro, 1997), such as those of EEMs. Parallel factor analysis (PARAFAC) is a multilinear model (Harshman, 1970; Leurgans and Ross, 1992; Bro, 1997; Andersen and Bro, 2003) suitable for such purpose. In fluorescence studies, PARAFAC has been widely utilized to perform spectral decomposition of EEMs and, thus, discriminate the profile of independent fluorophores (termed components) underlying the fluorescence signal of complex mixtures, such as DOM (Ohno and Bro, 2006).

Like PCA, PARAFAC decomposes the EEM three-way data array into sets of scores and loadings, condensing the multidimensional information into few underlying factors (components). Yet, it can be regarded as a higher order PCA, such that the model restricts each component to one score vector and two loading vectors (Bro, 1997). Thus, unlike PCA, a unique solution is given by providing the position of the axes defining a specific subspace (Harshman, 1970), which enables ease of interpretation.

The decomposition of the three-way array  $X$  of dimension  $I \times J \times K$ ,  $I$  referring to the number of samples,  $J$  to the emission spectra (mode) and  $K$  to the excitation spectra (mode) (Fig. 2.5), can be explained by the following equation:

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + \varepsilon_{ijk} \quad (2.22)$$

<sup>19</sup>The principle of parsimonious data modeling applied to multivariate calibration has been treated by Seasholtz and Kowalski (1993).

<sup>20</sup>A trilinear structure is defined by: (i) an equal number of components for each dimension (axis); (ii) each component has a unique profile for each dimension and (iii) its changes in one dimension are independent of any variation in the other two dimensions (i.e. the profile shape of a fluorophore derived from a specific combination of excitation and emission does not change varying its concentration) (Bro, 1997; Kumar and Mishra, 2013).

where  $x_{ijk}$  is the element of  $X$  representing the light intensity of the  $i$ th sample emitted at the  $j$ th variable when excited at the  $k$ th variable,  $F$  denotes the number of contributing factors (components),  $a_{if}$ ,  $b_{jf}$  and  $c_{kf}$  are the  $f$ th elements of the loading matrices  $A$ ,  $B$ ,  $C$  of dimensions  $I \times F$ ,  $J \times F$  and  $K \times F$ , and  $\varepsilon_{ijk}$  is the sum of squares of the residuals of dimension  $I \times J \times K$ . For fluorescence data, elements  $a_{if}$ ,  $b_{jf}$  and  $c_{kf}$  are a function of the relative contribution (i.e. concentration) of the mixture component  $f$  in the sample  $i$ , its emission ( $j$ ) and excitation ( $k$ ) spectra, respectively.

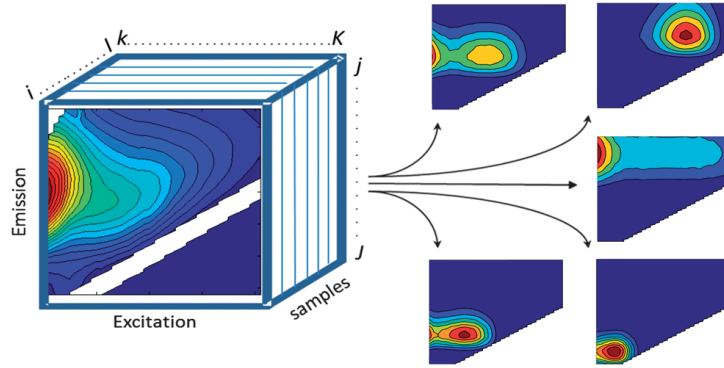
It follows that a simplified solution of the model can be expressed by the Khatri-Rao product (Bro, 1997):

$$X = \sum_{f=1}^F a_f \otimes b_f \otimes c_f \quad (2.23)$$

or alternatively (Burdick, 1995; Kumar, 2019):

$$X = A \times ((C^T \otimes B)^T)^T \quad (2.24)$$

where  $T$  indicates matrix transposition.



**Figure 2.5:** PARAFAC decomposition of the trilinear EEM structure in five components. Reprinted with permission from Murphy et al. (2013).

The PARAFAC model is commonly fitted with the alternate least square algorithm (Bro, 1997). The procedure involves the estimation of the loadings for two modes and then calculating the elements of the third mode and a residual array. This implies initialization of the spectral variables before PARAFAC computation, usually by singular value decomposition (SVD), which requires a relatively large number of iterations and considerable computational time. Yet, such a two-stage procedure allows an initial solution to be rotated towards a simple structure (Jolliffe, 2011) and the model fit to be improved at each consecutive iteration, until

convergence<sup>21</sup>. It is noteworthy that negative constraints are a fundamental criteria for achieving uniqueness of the solution, since neither concentration nor the fluorescence intensity can be negative.

*Pro et contra:* Such an approach can be regarded as a form of mathematical chromatography (Bro, 1997), enabling the non-destructive study of biochemical assemblies. As one of the most important features, uniqueness of the solution (Andersen and Bro, 2003) is given by rotational constraints, ensuring that the true underlying model is only the one with the best fit (Bro, 1997). Unlike other decomposition models (e.g. PCA), when adequately performed, PARAFAC uses less degrees of freedom, thus avoiding fitting potential noise (overfitting). Hence, while performing robust spectral decomposition in a structured, thus simpler manner, it is considered as the most parsimonious among other multilinear models. However, PARAFAC is not sequential; thus, the model is re-estimated when a different number of components is calculated. Advantageously, spectral components can be inferred regardless of interferences; nevertheless, consideration of the Raman and Rayleigh scattering, quenching and instrumental noise is important for an adequate model performance and output interpretation, since these commonly observed interferences can bias the model parameters (Andersen and Bro, 2003).

### Example: The MTT data set

PARAFAC was performed on EEMs obtained for the set of extracts from the MTT data set, previously described in this Subsection (2.5.3). Alternatively to the classical approach (Bro, 1997), the number of meaningful spectral components was determined by inferring their value in estimating a range of observed quantitative and qualitative WEOM properties using three regression algorithms. For each algorithm, the component performance was evaluated through: (i) the root mean squared error (RMSE), which scales off the errors to the same scale of the parameters and gives more weight to larger residuals than smaller ones; (ii) the  $R^2$  of the regression and (iii) the Pearson correlation coefficient ( $r$ ) for each variable analyzed. The PARAFAC results of this experiment are reported in Chapter 5 and related Appendix C.

---

<sup>21</sup>Iterations are generally stopped when the difference in the model fit between two consecutive iterations is  $10^{-6}$  (but also lower, e.g.  $10^{-10}$ ).

### 2.5.4 Machine learning techniques

#### Random Forest

The Random Forest (RF) algorithm constructs a collection of decorrelated decision trees<sup>22</sup> (non-linear models) and uses them to make a classification (Breiman, 2001) or regression. As a rule, about two-thirds of the original data set are used for the training purposes of the algorithm (called “training data set”), and about one-third for the validation of the predictive power of the trained algorithm (“out-of-bag data set”), as well as the estimation of its error (generalization error).

Here is an example of the RF algorithm logic:

$$S = \begin{bmatrix} f_{A1} & f_{B1} & f_{C1} & C_1 \\ \vdots & & \vdots & \\ f_{An} & f_{Bn} & f_{Cn} & C_n \end{bmatrix} \quad (2.25)$$

where  $S$  represents a matrix of training samples utilized by the algorithm to build a classification model,  $A1$  and  $An$  represent the feature  $A$  of the first and the  $n$ -th sample, and  $C_1...C_n$  indicate the levels or categories of each sample in the bootstrapped matrix. The algorithm creates a number of subsets (i.e.  $S_1...S_n$ ) with randomly selected features (variables)<sup>23</sup> (known as “the random subspace method”), each of which is used to split a node essential to grow a decision tree from the correspondent bootstrapped training data set. Finally, the large number of decision trees generated (forest) is used to create a ranking of classifiers for prediction, based on the number of votes (list of probabilities) collected for each class (classification), or for regression, in which the sum of the predicted numeric values from the forest is averaged by the number of trees. Generally, the bigger the forest, the more robust the prediction, the higher the accuracy.

The outputs of such computation are two measures of variable importance: the Increased Mean Square Error (IncMSE) and Increased Node Purity Index (IncNodePurity). The IncMSE estimates the increase in the mean squared error when values for a specific variable are randomly permuted without altering the other variables (Breiman, 2001). The importance of the variable increases if the random permutation changes the predicted value. The IncNodePurity measures the mean reduction in the residual sum of squares (increase in node purity or homogeneity) resulting from splitting the data on a given variable.

<sup>22</sup>Successive trees are independently constructed using a bootstrap sample of the original data set and do not depend on previously generated trees, a method called *bagging* (Breiman, 1996).

<sup>23</sup>Typically, the optimal number of features is selected comparing the generalization error of the predictive algorithm using the square root of the number of variables and a couple of settings above and below that number.

*Pro et contra:* Such an approach favors both classification and regression outcomes, especially when the number of predictors exceeds the sample size (the so-called “ $p > n$  problem”), it does not require assumptions based on data distribution, and avoids overfitting<sup>22</sup> that might occur as a result of higher-order interactions between variable inputs (Breiman, 2001; Lagomarsino et al., 2017). Yet, predictions are not made beyond the range of training data and overfitting may still occur for particularly noisy data sets.

These attributes make RF one of the most powerful and robust supervised<sup>24</sup> machine learning techniques, especially for the analysis of complex, non-linear and high-dimensional data (Lagomarsino et al., 2017).

### Example: The MTT data set

An example of the RF implementation in this study is presented in Chapter 4. Training and cross-validation (10-fold) for model selection were done in R via the ‘caret’ package (Kuhn, 2008, 2015). The RF was then run using the ‘randomForest’ package (Liaw and Wiener, 2002), as shown in the following example:

```
> xtrain <- dataTbl[, 8:17]
> rf_fit <- randomForest(CO2 ~., data = xtrain, importance = TRUE,
                        proximity = TRUE, ntree = 800, mtry = 9)

> print(rf_fit)

Type of random forest:  regression
Number of trees: 800
No. of variables tried at each split: 9

Mean of squared residuals: 1891.307
% Var explained: 75.81
```

---

<sup>24</sup>In supervised machine learning, the algorithm uses the training data set to learn the mapping function from the input variables (x) to the output variable (Y) and make predictions.

This page is intentionally left blank

## Part I

# Behind the scenes of DOM biodegradability

This page is intentionally left blank



## Chapter 3

### Quantity and biodegradability of DOM released from sequentially leached soils, as influenced by the extent of soil drying prior to rewetting

This chapter has been published in the journal Soil Research and adapted with the approval of all co-authors.

Tihana Vujinović, Timothy J. Clough, Denis Curtin, Esther D. Meenken,  
Niklas J. Lehto, and Michael H. Beare

Soil Research, 2019  
Volume 57(4), 374–386

## Abstract

Soil rewetting can induce a flush of organic matter mineralization, but the factors underpinning this mineralization response are poorly understood. We investigated the effects of antecedent soil water content, before rewetting, on the quantity, quality and biodegradability of dissolved organic matter present in the leachate pore volumes from a soil under two different management histories: arable and grassland. Soils were collected at field capacity (FC) and dried to give four soil gravimetric water contents ( $\theta_g$ ): 22% (not dried, left at FC), 15%, 8% and <2% (air dry, AD). Soils were repacked to the same bulk density ( $1.1 \text{ g cm}^{-3}$ ) and each core was sequentially leached, with four pore volumes collected. The total amount of dissolved organic carbon (DOC) leached increased ( $P < 0.001$ ) only in the soils that had been air-dried before rewetting (3.8 and  $5.3 \text{ mg g}^{-1}$  soil C, for arable and grassland, respectively), while among the other  $\theta_g$  treatments differences were relatively small ( $1.6\text{--}2.4 \text{ mg g}^{-1}$  soil C). The pre-rewetting  $\theta_g$  treatment affected the DOC content of the pore volume leached ( $P < 0.001$ ): in the grassland soil, the DOC of the AD treatment was consistently twice as high as the other  $\theta_g$  treatments, but this trend was not as consistent in the arable soil. For all  $\theta_g$  treatments and both soils, specific ultraviolet absorbance at 254 nm increased as leaching progressed. Biodegradability, expressed as cumulative  $\text{CO}_2$  produced per unit of DOC in leachates, was significantly lower in the first pore volume of all treatments in the grassland soil and increased with sequential leaching. In the arable soil, differences were small or insignificant across the pore volumes leached, but were large and inconsistent across the  $\theta_g$  treatments. These findings improve our understanding of how antecedent soil water content affects the quantity and quality of dissolved organic matter released when soils are rewetted, and the potential for soil carbon losses.

**Keywords:** Arable, Dissolved organic carbon, Biodegradation, Drying and rewetting, Grassland,  $\text{SUVA}_{254}$

## 3.1 Introduction

Dissolved organic matter (DOM) in soil solution is recognized as the most mobile (von Lützow et al., 2007) and reactive (Haynes, 2005) soil organic matter (SOM) fraction, and its turnover thus represents a crucial step in carbon (C) and nitrogen (N) cycling (Kalbitz et al., 2000; Zsolnay, 2003). The physico-chemical interaction of DOM with the soil matrix (Stevenson, 1994) can hinder microbial accessibility (Sollins et al., 1996; Hodson et al., 2011) to potentially bioavailable DOM (Boyer and Groffman, 1996; Marschner and Kalbitz, 2003; Haynes, 2005). Furthermore, as

it comprises a highly complex and heterogeneous mixture of organic compounds, not all of the bioavailable DOM is readily biodegradable: its turnover times may span hours to years (Marschner and Kalbitz, 2003). Thus, both the release of DOM and its potential to be mineralized, hereafter referred to as biodegradability, have a direct bearing on the transformations and fate of C and N.

The degree of DOM biodegradability is related to its composition. The DOM consists of rapidly degradable components, such as simple carbohydrates, proteins, low molecular weight organic acids, amino acids, amino sugars, fatty acids and a large pool of non-specified organic compounds (Qualls and Haines, 1992; Guggenberger et al., 1994b; Gregorich et al., 2003; Poirier et al., 2005), but also contains lignin fragments, complex polysaccharides and microbially derived polymeric compounds (Kalbitz et al., 2000), which are degradable only with a specific set of enzymes. In spite of this, Kalbitz et al. (2003) found DOM biodegradability to be closely, but non-linearly, related to DOM's chemical and structural properties (i.e. DOM quality). Their study, recently expanded upon by Apostel et al. (2015), confirmed the role of carbohydrates in the first step of microbial utilization, while corroborating the concept that low mineralization rate constants are associated with aromatic moieties and double-bonded alkyl compounds (Baldock et al., 1992; Qualls and Haines, 1992; Bolan et al., 2011). However, mineralization of aromatic hydrocarbons has also been demonstrated (Scow et al., 1989; Stapleton et al., 1998; Almendros and Dorado, 1999). Parton et al. (2015) argued that breakdown occurs when the aromatic substrate is available; however, the extent to which this occurs is driven by complex interactions between the microbial communities, the degree of oxidative alteration, the internal cross-linking of aromatic moieties and nutrient cofactors (Scow et al., 1989; Guggenberger and Zech, 1994; Stapleton et al., 1998; Almendros and Dorado, 1999), which often leads to disparate correlations and contrasting conclusions. In addition to DOM quality, Marschner and Kalbitz (2003) also stressed that biodegradation is a function of external factors (such as temperature, rainfall regime and associated vegetational cycles), which can directly influence DOM's chemical characteristics, as well as soil microbial activity (Sinsabaugh and Follstad Shah, 2012).

In unsaturated soil systems, soil water is a key factor regulating DOM release (Zsolnay et al., 1999) and C mineralization (Chow et al., 2006). Although it is well established that soil microbial activity and community structure play a fundamental role in OM solubilization (Kuzyakov et al., 2000; Cheng and Kuzyakov, 2005) and DOM mineralization, under field conditions, abiotic factors, such as fluctuations in soil moisture, may play a larger role in the very first steps of OM cycling (Kemmitt et al., 2008) by promoting both abiotic oxidation of DOM (Wu and Brookes, 2005) and affecting microbial metabolism and community structure (Fierer et al., 2003). Drying–rewetting events are known to enhance the release (Fierer and Schimel,

2002; Harrison-Kirk et al., 2014) and mineralization of DOM (Birch, 1958). Although a positive correlation has been found between DOC concentration and C mineralization with dry–wet cycles, the magnitude of the mineralization response can differ even when DOC concentrations are similar (Chow et al., 2006; Harrison-Kirk et al., 2014). This suggests that changes in DOM quality (affected by fluctuations in soil moisture) may have a greater impact on the magnitude of the DOM mineralization response compared to the size of this pool. Hence, uncertainties about the relationship between the availability, biodegradability and composition of DOM in this scenario urge further examination.

Most studies on soil DOM quantity and quality have been based on soil extracts, obtained with different extraction protocols. This implies that DOM characteristics may be affected by pre-extraction soil treatments (e.g. air-drying) and extraction parameters, such as temperature, soil:extractant ratio, extraction time, pH and ionic concentration of the extractant (Zsolnay, 2003; Guigue et al., 2014). As a result, extracted DOM may not be correlated with DOM release *in situ* (DOM in soil leachates). Furthermore, bulk extractions cannot provide information on the dynamics of DOM release. Studies on DOM derived from bulk leachate samples have been performed (Ghani et al., 2010) but, to our knowledge, none have examined the potential of DOM quality to change as volumes of the soil solution are progressively leached out.

Thus, the objectives of this study were to (i) assess how the quantity, quality and biodegradability of DOM responded to different degrees of soil drying before rewetting and (ii) determine how these measures of DOM varied over the course of a leaching event.

## 3.2 Materials and methods

### 3.2.1 Soil sampling and characterization

Soil samples were collected (0–15 cm depth) from a long-term arable site and from a perennial grassland site (‘Grasslands Nui’ perennial ryegrass (*Lolium perenne* L.) and white clover (*Trifolium repens* L.)) at Lincoln, Canterbury, New Zealand (43°38’ S, 172°28’ E) in April 2016. The arable site had been cultivated and rotationally cropped (approximately every 3 years, with a crop rotation of barley (*Hordeum vulgare* L.) – ryegrass (*Lolium* spp.) – wheat (*Triticum* spp.) – pea (*Pisum sativum* L.)) for the previous 18 years. Both sampling sites were within 200 m of each other, on a Templeton silt-loam soil (classified as Eutric Siltic Cambisol, IUSS Working Group WRB, 2015). The sites had not been fertilized for at least three months before sampling. Three randomly selected samples (~5 kg) were taken from a 1 m<sup>2</sup> area within each site that had been previously irrigated to bring the soils to field

capacity.

In the laboratory, the field-moist soils were sieved ( $<4$  mm) and coarse plant residues and visible fauna removed by hand. Total organic C and N were measured on a LECO TruMac CNS analyzer (LECO Corporation, Saint Joseph, MI, USA). Soil texture, bulk density and pH in water were determined following standard protocols (Gee and Or, 2002; Carter and Gregorich, 2008; Devey et al., 2010). Mineral N was determined by extracting soils with 2 M KCl for 1 h at a soil:extractant ratio of 1:5. The filtered extracts were colorimetrically analyzed for ammonium ( $\text{NH}_4^+$ )-N and nitrate ( $\text{NO}_3^-$ )-N on a Lachat QuickChem 8500 Series 2 Flow Injection Analysis System (Lachat Instruments, Loveland, CO, USA) (Keeney and Nelson, 1982). Soil properties at each site are given in Table 3.1.

**Table 3.1:** Physical and chemical properties of arable and grassland soils.

Land use	Sand (%)	Silt (%)	Clay (%)	pH (H <sub>2</sub> O)	TOC (g kg <sup>-1</sup> )	TN (g kg <sup>-1</sup> )	Min N (mg kg <sup>-1</sup> )	$\theta_g$ (%)
Arable	34 (1.7)	46 (2.3)	15 (0.8)	4.7	20.1 $\pm$ 0.88	1.8 $\pm$ 0.04	64.1 $\pm$ 0.10	21.2 $\pm$ 0.04
Grassland	33 (1.7)	43 (2.2)	18 (0.9)	5.6	29.8 $\pm$ 0.17	2.4 $\pm$ 0.02	10.5 $\pm$ 0.15	22.6 $\pm$ 0.06

TOC, total organic carbon ( $n = 2$ ); TN, total nitrogen ( $n = 2$ ); Min N, mineral nitrogen (nitrate + ammonium;  $n = 6$ );  $\theta_g$ , gravimetric water content at field capacity ( $n = 3$ ). Soil textural data are reported with 5% error (in brackets). Values are mean  $\pm$  s.e.m.

### 3.2.2 Experimental treatments

Treatments were established to give four gravimetric water contents ( $\theta_g$ ): 22% (field capacity), 15%, 8% and  $<2\%$  (air dry) – hereafter referred to as FC, 15%, 8% and AD, respectively. The  $\theta_g$  treatments were established by drying the soil from its initial FC moisture content at low temperature (4–6°C) and constant humidity (this allowed controlled drying, while slowing down microbial activity). Once the designated moisture contents were reached and verified, by determining weight loss in subsamples following oven drying at 105°C, soils were stored at 4°C in sealed glass preserving jars for no more than 72 h before repacking and rewetting.

Total porosity ( $\varepsilon$ ), air-filled porosity ( $\varepsilon_a$ ) and water-filled pore space of the repacked cores (Table 3.2) were calculated (Mclaren and Cameron, 1996). One pore volume (PV) was defined as  $\varepsilon$  multiplied by the volume of the repacked core. The matric potential ( $\psi_m$ ) of each  $\theta_g$  treatment was determined by fitting the volumetric

water contents ( $\theta_v$ ) at  $1.1 \text{ g cm}^{-3}$ , determined in pressure plates on disturbed core samples at suction pressures of 30, 100, 200, 500, 1000 and 1500 kPa, to the van Genuchten (1980) equation. The parameters of the van Genuchten model were estimated using the RETC program (van Genuchten et al., 1991) and  $\theta_s$  (saturated water content) was set equal to  $\varepsilon$ . Finally, the maximum diameter of water-filled pores ( $d$ , McLaren and Cameron, 1996) was estimated for each  $\theta_g$  treatment.

**Table 3.2:** Physical properties of repacked arable and grassland soil cores.

Land use	Variable	$\theta_g$ treatment			
		FC	15%	8%	AD
Arable	$\theta_v (\text{cm}^3 \text{ cm}^{-3})$	0.23	0.17	0.09	0.02
	$\varepsilon_a (\text{cm}^3 \text{ cm}^{-3})$	0.35	0.41	0.49	0.56
	WFPS (%)	39.9	29.7	15.6	3.6
	$\psi_m (\text{MPa})$	0.01	0.03	0.4	144.8
	$d (\mu\text{m})$	30.2	9.4	0.75	0.002
Grassland	$\theta_v (\text{cm}^3 \text{ cm}^{-3})$	0.25	0.16	0.08	0.03
	$\varepsilon_a (\text{cm}^3 \text{ cm}^{-3})$	0.34	0.42	0.51	0.56
	WFPS (%)	42.4	27.4	13.5	4.8
	$\psi_m (\text{MPa})$	0.01	0.05	0.9	61.5
	$d (\mu\text{m})$	31.2	5.5	0.32	0.005

Bulk density =  $1.1 \text{ g cm}^{-3}$ ; total porosity = 58%; pore volume = 34.7 mL. Results are given as a mean ( $n = 3$ ) expressed on an oven-dry weight basis.

Abbreviations:  $\theta_g$ , gravimetric water content;  $\theta_v$ , volumetric water content;  $\varepsilon_a$ , air-filled porosity; WFPS, water-filled pore space;  $\psi_m$ , soil matric potential, as back-calculated values predicted from the van Genuchten (1980) model;  $d$ , maximum pore diameter full of water.

### 3.2.3 Leachate collection

Arable and grassland soils were repacked (at their designated  $\theta_g$ ) into plastic Buchner funnel tops ( $\varnothing$  5.5 cm,  $h = 3$  cm) to a depth of 2.5 cm, with a dry bulk density ( $\rho_b$ ) of  $1.1 \text{ g cm}^{-3}$ . Glass microfiber filter paper (Whatman GF/A,  $\varnothing$  55 mm) was placed at the base of the funnel tops to prevent the loss of fine particles during the

leaching process. Each  $\theta_g$  treatment was replicated six times. Batches of repacked soil cores at each  $\theta_g$  were slowly wetted with deionized water to FC and allowed to equilibrate for  $\sim 1$  h. A glass microfiber filter (Whatman GF/A,  $\varnothing$  55 mm) was placed on each core to ensure even distribution of water over the entire surface and to prevent structural disturbance.

Each core was then leached with four PVs (one PV = 35 mL) of deionized water, by dripping water onto the soil surface at a constant rate. This was done using a leaching manifold equipped with sterile plastic syringes (fitted with 25-gauge needles), enabling 12 soil cores to be leached simultaneously. The PVs of leachate were collected, after  $\sim 1$ –2 h, in separate 50-mL Falcon<sup>™</sup> tubes and centrifuged at 4654 *g* for 10 min before filtration through a 0.45- $\mu$ m polyvinylidene fluoride (PVDF) syringe filter. All leachates were analyzed within 24 h and stored at 4°C overnight before subsequent incubation.

### 3.2.4 Leachate sample analyses and characterization

Leachates were analyzed for total organic carbon (TOC) concentrations, determined as non-purgeable organic carbon (NPOC), on a Shimadzu TOC- $V_{CSH}$  analyzer (Shimadzu Corporation, Kyoto, Japan) fitted with a high-sensitivity catalyst for low TOC concentrations. Instrumental drift was checked by running deionized water and a 20 mg C L<sup>-1</sup> standard solution over the course of the analyses. Total dissolved nitrogen (TDN) was analyzed using an on-line UV-catalyzed persulfate oxidation method (Cabrera and Beare, 1993) on a Lachat QuickChem 8500 Series 2 Flow Injection Analysis System; leachate DON concentrations were then calculated by subtracting the inorganic N forms (NO<sub>3</sub><sup>-</sup>-N and NH<sub>4</sub><sup>+</sup>-N) from TDN.

The UV absorbance at 254 nm of diluted (10-fold) leachates was determined on a UV-Vis SpectraMax 190 microplate reader (Molecular Devices Corp., San Jose, CA, USA) and normalized for the total DOC concentration. These data were recorded with units of L mg<sup>-1</sup> m<sup>-1</sup> of specific UV absorbance (SUVA<sub>254</sub>), which has previously been shown to be positively correlated with the aromaticity of DOM (Novak et al., 1992; Weishaar et al., 2003; Jaffrain et al., 2007), as well as with conjugated unsaturated C heteroatoms, such as olefins (Weishaar et al., 2003).

Leachate pH was measured directly with a calibrated Mettler Toledo Seven Easy pH-meter (Mettler Toledo, Urdorf, Switzerland).

### 3.2.5 Water extractable organic C (WEOC)

Five replicates of each soil, at their designated  $\theta_g$ , were extracted with deionized water by shaking for 1 h at a soil:water ratio (dry weight basis) of 1:10. The extracts were then centrifuged at 2968 *g* for 10 min and the supernatants were re-centrifuged

at 11 872 *g* for 5 min prior to filtration (0.45- $\mu$ m PVDF syringe filters). Extracts were analyzed for NPOC within 24 h.

### 3.2.6 Biodegradation assay

The biodegradability of DOM in the leachates was determined using a 14-day static incubation assay, adapted from those described by Gregorich et al. (2003) and Kalbitz et al. (2003). Twenty milliliters of each leachate, 50  $\mu$ L of unfiltered microbial inoculum and 1 mL of an N-free nutrient solution (to avoid nutrient limitations to DOM decomposition) were incubated at 20°C for 14 days in sealed 100-mL Schott bottles with rubber septa inserted in the lid to allow for headspace gas sampling. Samples were assayed in quintuplicate. Positive and negative controls were run in triplicate for each batch incubation. The positive control contained 20 mL of glucose solution (20 mg C L<sup>-1</sup>) in place of the leachate, while the negative control contained 20 mL of deionized water. The inoculum (2.1 mg C g<sup>-1</sup> soil) was prepared from a combined sample of both soils. Briefly, 4 g of soil were shaken with 45 mL of deionized water before incubation at 35°C for 24 h. The inoculum was then re-suspended and allowed to stand for  $\sim$ 1 h before transferring the upper 30 mL into a clean 50-mL Falcon<sup>™</sup> tube.

The amount of DOC mineralized (DOC-C<sub>min</sub>) was assessed as cumulative CO<sub>2</sub> produced over a period of 14 days. Gas samples were collected on days 1, 2, 4, 7, 10 and 14 of the incubation using a 20-mL gas-tight syringe after thoroughly mixing the headspace, by pumping the syringe several times before withdrawing the sample. Sample CO<sub>2</sub> concentrations were determined by directly injecting the gas samples into a Li-7000 infrared gas analyzer (IRGA, LI-COR Inc., Lincoln, NE, USA). Following each sampling, the incubation bottles were opened and flushed to re-establish an atmospheric CO<sub>2</sub> concentration before being returned to the incubator. For each  $\theta_g$  treatment, CO<sub>2</sub> fluxes were integrated over time, in order to calculate the cumulative CO<sub>2</sub> evolved. Results were expressed as CO<sub>2</sub>-C produced per unit of DOC present in the original leachate (Baldock and Broos, 2012).

### 3.2.7 Statistical analysis

Treatment effects on the characteristics of DOM leached – namely DOC quantity (mg g<sup>-1</sup> soil C), DOC-C<sub>min</sub> (mg CO<sub>2</sub>-C g<sup>-1</sup> DOC), DON and mineral N (mg g<sup>-1</sup> soil N), SUVA<sub>254</sub> (L mg<sup>-1</sup> m<sup>-1</sup>) and pH – from each of the four sequential soil PVs were evaluated using a linear mixed model fitted with Restricted Maximum Likelihood (REML). Variability was expressed by the least significant differences (l.s.d.) at the 5% level of confidence.

The relationships between the quantitative and qualitative variables and the leachate DOC mineralized were investigated using a multivariate regression analysis.



The PVs were always clustered regardless of soil or  $\theta_g$  treatment; thus, this factor was not included in the regression analysis.

During both the regression and REML model fitting, the relationships between the response and each predictive parameter by  $\theta_g$  treatment were quite different. For this reason, the  $\theta_g$  factor was partitioned to create a series of nested terms (orthogonal contrasts) to help formalize how soil moisture affected the relationships between quantitative and qualitative variables. This REML modeling proceeded as follows: first, the overall response of the AD treatment was compared with the mean of the remaining  $\theta_g$  treatments; second, the other extreme treatment, FC, was compared with the mean of the intermediate  $\theta_g$  treatments; finally, the remaining  $\theta_g$  treatments were compared. For the regression model fitting, only the first (AD vs the mean of the other  $\theta_g$  treatments) and final partitions (at least one of the other  $\theta_g$  treatments differed significantly) were tested. The model that was most parsimonious was selected via backward selection.

Each soil was analyzed separately because of the lack of replication for the land use factor, as discussed in the methods. All statistical analyses were performed using the Genstat statistical package version 17.0. Graphs were prepared using SigmaPlot version 12.5.

### 3.3 Results and discussion

#### 3.3.1 Soil properties

The two sites had similar soil texture, but differed in organic matter content; the grassland soil had  $\sim 50\%$  more C and  $\sim 30\%$  more N than the arable soil (Table 3.1). The C:N ratios in the arable and grassland soils were 11:1 and 12:1, respectively. Despite the difference in organic matter content,  $\theta_g$  at FC was similar for both soils ( $\sim 22\%$ ). The arable soil had a lower pH and a higher mineral N content (dominated by  $\text{NO}_3^-$ -N) than the grassland soil, presumably due to the nitrification of  $\text{NH}_4^+$ -based fertilizers utilized in intensive agriculture. The physical properties of the repacked soil cores at the corresponding volumetric water content ( $\theta_v$ ) for each  $\theta_g$  treatment are shown in Table 3.2. As expected,  $\varepsilon_a$  increased and water-filled pore space (WFPS) decreased with decreasing  $\theta_g$ . At FC, the diameter of the pores (macropores) filled with water was consistent for both soils, but with increasing soil drying (mesopores) it was about half the value in the grassland soil (Table 3.2).

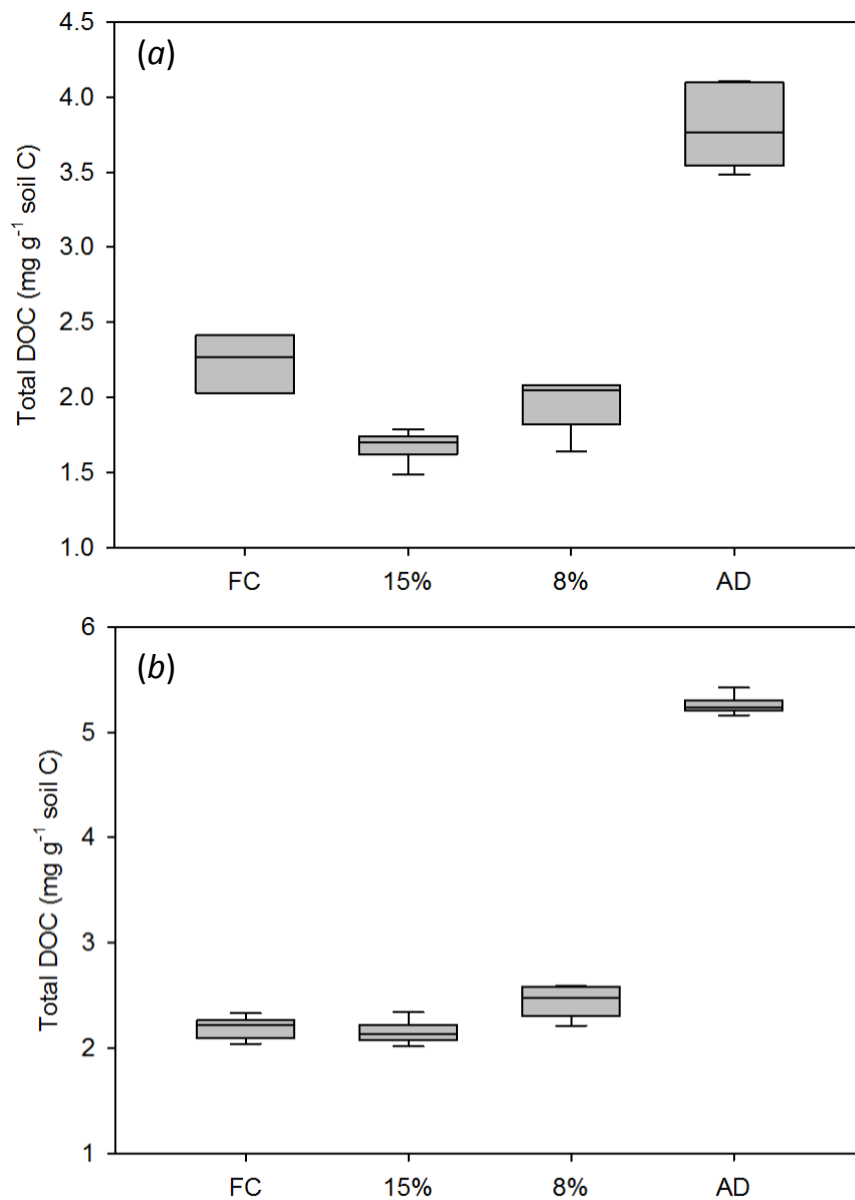
#### 3.3.2 DOC in soil leachates

The total DOC leached (sum of DOC leached over four PVs) was influenced by the  $\theta_g$  treatment ( $P < 0.001$ ), with more DOC leached from the AD treatment

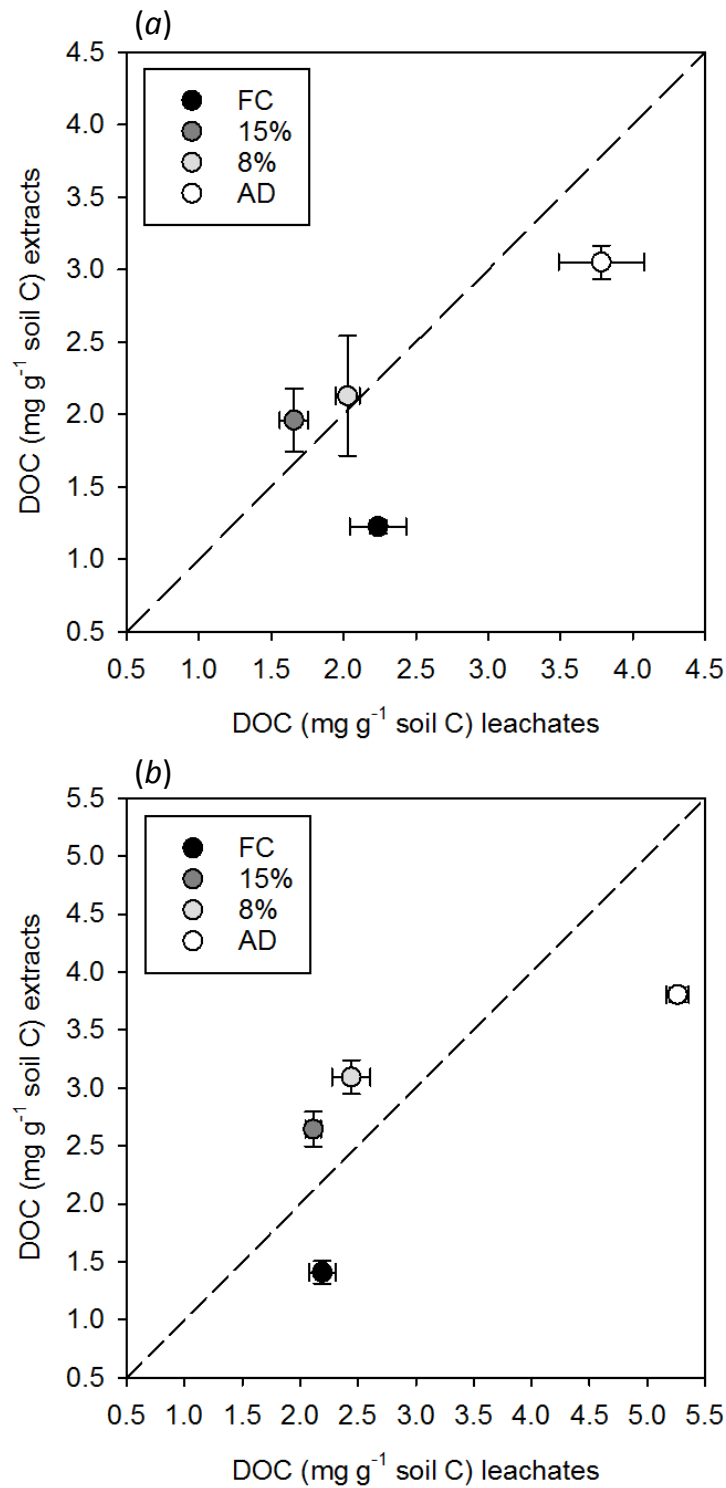
(Fig. 3.1). Among the other  $\theta_g$  treatments, differences in the total amount of DOC leached were relatively small. This is consistent with previous studies (Ludwig et al., 1999; Kaiser et al., 2001) that reported WEOC to be 2–5 times higher in air-dried soil samples than in field-moist soils (regardless of land use history). In the current experiment, the WEOC and total DOC leached were comparable for the soils when dried to 15% and 8%  $\theta_g$  (Fig. 3.2); however, for the FC and AD treatments, the total DOC leached was greater than that extracted. Reasons for this variation are not obvious. It is possible that the time required to collect a PV caused the observed variation between WEOC and DOC leached. However, for both soils, mean times to collect PVs at 15% and 8%  $\theta_g$  were both 1.5 h, and this was also the case in the AD treatment. It was only in the FC treatment of the arable soil where the mean PV collection time was greater (2.5 h). Thus, changes in PV collection time in response to the drying treatments are inconsistent with the variation seen between extractable and total DOC leached.

Comparing successive PVs in the arable soil (Fig. 3.3a) showed that the rate of change in the amount of DOC leached varied with  $\theta_g$  treatment: the DOC decreased from the first to the fourth PV in the AD and 8% treatments (47% and 42%, respectively); increased (up to 61%) from the first to the fourth PV in the FC treatment, despite a significant decrease between the second and third PV; but did not differ in the 15% treatment (l.s.d. > 0.05) until the fourth PV leached (27% decrease). In the grassland soil (Fig. 3.3b), the DOC concentration decreased systematically from the first to the third PV of all  $\theta_g$  treatments (l.s.d. < 0.05), with a further decrease from the third to the fourth PV in the AD treatment only. The rate of decrease in DOC concentration was relatively constant for all  $\theta_g$  treatments (63–68%). The overall decrease in the content of DOC across successive PVs (apart from the arable FC treatment) was in line with the assumption that continued leaching would progressively remove DOC in a dynamic system without external OM inputs.

However, the extent of soil drying before rewetting affected the DOC concentration within the PV leached ( $P < 0.001$ , Fig. 3.3). Regardless of land use, the PVs of the AD treatment had the highest DOC concentration. Clearly, the mechanism(s) responsible for enhancing the release of DOC, regardless of PV and land use, is(are) triggered to a greater extent only in AD soils. It is well recognized that aggregate disruption enhances the release of DOC following air-drying as a consequence of air compression, generated by surface tension occurring during the aggregate rewetting, exceeding the cohesion between soil mineral particles and aggregates (Kemper et al., 1987), as well as the uneven hydration and swelling of the clay fraction (Caron et al., 1996). Furthermore, we cannot exclude the breakdown of microbial cells (Bottner, 1985; Van Gestel et al., 1993a) and the release of intracellular solutes (Halverson et al., 2000) induced by an increased osmotic pressure,



**Figure 3.1:** Boxplots showing total dissolved organic C (sum of DOC in four pore volumes) leached from an arable (a) and a grassland soil (b) at varying degrees of  $\theta_g$ : field capacity (FC), 15%, 8% and <2%  $\theta_g$  (air dry; AD). The upper and lower boundaries of the boxes show the 25<sup>th</sup> and 75<sup>th</sup> percentile, respectively; the central line represents the median; the whiskers indicate the minimum and maximum values. For the arable soil at FC  $n = 3$ , for the other  $\theta_g$  treatments of both soils  $n = 6$ .



**Figure 3.2:** The relationship between total dissolved organic C (DOC; mg g<sup>-1</sup> soil C) of leachates (for the arable soil field capacity (FC) treatment  $n = 3$ , for the other  $\theta_g$  treatments  $n = 6$ ) and extracts ( $n = 5$ ) from an arable (a) and a grassland soil (b) at FC, 15%  $\theta_g$ , 8%  $\theta_g$  and air dry (AD). The dashed line represents a hypothetical 1:1 ratio. Bars indicate the standard deviation.

brought about by complete dehydration, as possible mechanisms.

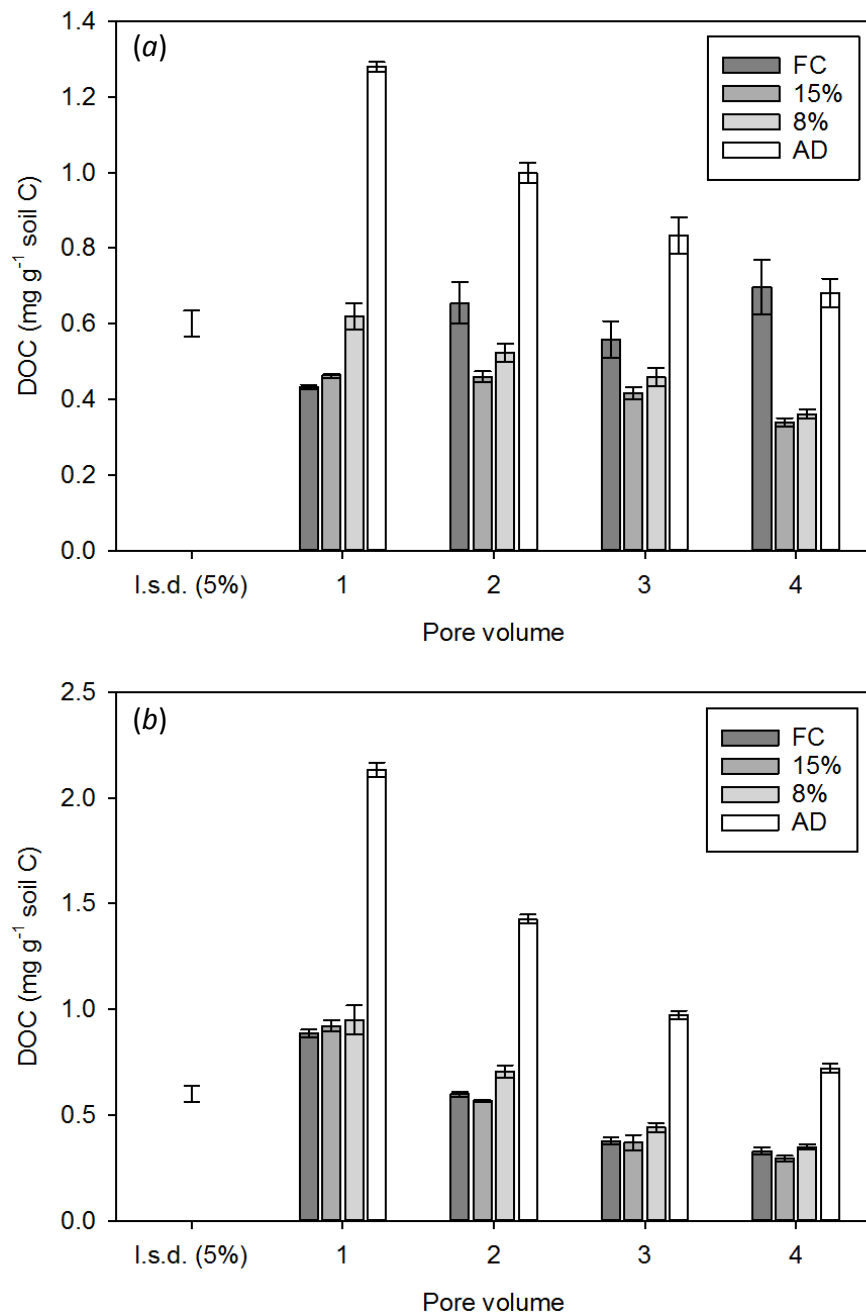
Relatively high levels of SOM found in grassland and no-tillage soils promote aggregate stability (Tisdall et al., 1978; Haynes and Swift, 1990; Cambardella and Elliott, 1993; Six et al., 2004) and decrease aggregate wettability (Caron et al., 1996). Assuming such an effect in the grassland soil, this may explain the lack of observed differences between the FC, 15% and 8% treatments, with respect to PV DOC concentrations (Fig. 3.3*b*).

It is generally recognized that arable soils have low SOM content and, therefore, they commonly manifest lower aggregate stability (Paustian et al., 2000; Bronick and Lal, 2005; Vogelmann et al., 2013). Assuming that this was the case in the arable soil, this may explain the variability in the PV DOC concentrations for the FC, 15% and 8% treatments (Fig. 3.3*a*). Contributing to this variability was the decrease in the rate of leaching in the FC treatment (from 2 to 3 h for PVs 2 to 4), compared with the AD treatment of the arable soil (1–1.5 h for all PVs), that may have increased the contact time between the soil and the water phase, thus leading to more DOC being released. However, these postulated effects cannot be directly confirmed without more intensive measurements of aggregate stability and contact time effects.

Although our study does not allow for a robust comparison of land use effects, it shows that approximately twice as much DOC per unit of soil C was released with the first PV of the grassland soil relative to the arable soil. This suggests either a greater soluble C pool or a greater efficiency of soluble C extraction by infiltrating water. Dexter (1988) reported that macroaggregates tend to break apart more easily than microaggregates when wetted up because of their larger macroporosity, and it is recognized that grassland soils contain a larger fraction of macroaggregates than arable soils (Horwath, 2015). If, indeed, this was the case, macroaggregate disruption could have led to the release of inter-aggregate soluble C (Golchin et al., 1994; Miller et al., 2005; Smucker et al., 2007). Another explanation would be a reduced content of intra-aggregate and/or mineral-associated soluble C in the arable soil, as a result of cultivation (Cambardella and Elliott, 1992; Grandy and Robertson, 2006).

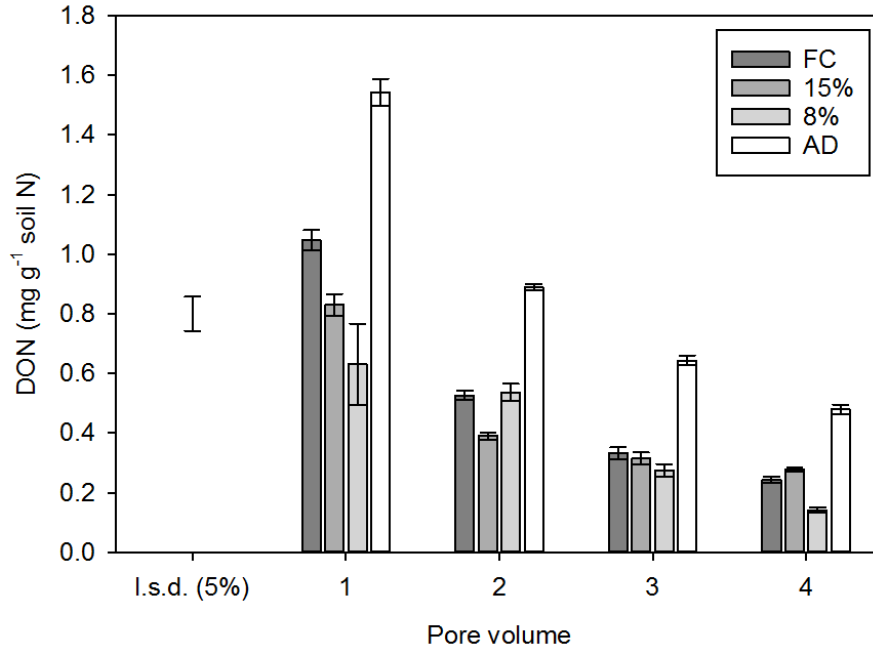
### 3.3.3 Leachate N

Leachate DON represented only a small portion of the TDN in the first PV of both soils, as TDN in the first PV was dominated by inorganic N ( $\text{NO}_3^-$ -N) (up to 98% in the arable and 88% in the grassland soil, data not shown). As the arable soil released up to five times more TDN in the first PV than the grassland soil (data not shown), a cumulative effect on the error associated with the  $\text{NO}_3^-$ -N proportion of the TDN in the arable soil might have strongly biased the calculated DON concentration. Thus, only data on DON released from the grassland soil are



**Figure 3.3:** Dissolved organic C content (DOC; mean  $\pm$  s.e.m.) in the pore volumes leached from an arable (a) and a grassland soil (b) at four levels of  $\theta_g$ . l.s.d. ( $\alpha = 0.05$ ) = 0.07 and 0.08  $\text{mg g}^{-1}$  soil C for arable and grassland, respectively. FC, field capacity; AD, air dry.

presented (Fig. 3.4).



**Figure 3.4:** Dissolved organic N content (DON; mean  $\pm$  s.e.m.) in the pore volumes leached from a grassland soil at four levels of  $\theta_g$ . l.s.d. ( $\alpha = 0.05$ ) = 0.12 mg g<sup>-1</sup> soil N. FC, field capacity; AD, air dry.

The DON leachate concentration relationship to PV varied with  $\theta_g$  treatment ( $P < 0.001$  for the interaction of PV with  $\theta_g$  treatment). Most of the DON was released in the first PV leached (40–49% of the total), and this gradually decreased by  $\sim 30\%$  (AD) to 37% (FC) in the fourth PV. In the first PV, more DON was released from the AD treatment than from the other  $\theta_g$  treatments, reflecting the effect of air-drying previously described for the DOC release. However, the FC treatment leached 18–36% more DON with respect to the 15% and 8% treatments ( $P < 0.001$ ). We observed the same recurrent effect of the AD treatment in all the subsequent PVs leached.

### 3.3.4 SUVA<sub>254</sub>

Ultraviolet light absorption at 254 nm per unit of C (SUVA<sub>254</sub>) has been shown to be directly proportional to the aromatic content (as well as to conjugated aliphatic moieties, such as olefins) of DOM (Weishaar et al., 2003). Leachate SUVA<sub>254</sub> was influenced by the  $\theta_g$  treatment and the PV leached (Table 3.3) with, generally, higher SUVA<sub>254</sub> values ( $P < 0.001$ ) in the wetter soil treatments (FC and 15%) than

in the drier treatments (8% and AD). These findings are consistent with the results reported by Chow et al. (2006), who found higher SUVA<sub>254</sub> in wetter treatments of an agricultural peat surface soil with respect to the drier counterparts.

**Table 3.3:** SUVA<sub>254</sub> values (L mg<sup>-1</sup> m<sup>-1</sup>) in the pore volumes leached from an arable and a grassland soil at their  $\theta_g$  treatment before rewetting.

Land use	Pore volume	$\theta_g$ treatment			
		FC	15%	8%	AD
Arable	1	2.86 ± 0.06	4.08 ± 0.17	2.95 ± 0.11	2.40 ± 0.13
	2	10.12 ± 0.47	8.23 ± 0.37	4.96 ± 0.38	3.52 ± 0.10
	3	10.00 ± 0.59	11.31 ± 0.45	5.46 ± 0.22	4.06 ± 0.06
	4	7.48 ± 1.02	13.01 ± 0.52	6.04 ± 0.26	4.50 ± 0.07
Grassland	1	5.80 ± 0.21	4.28 ± 0.05	3.73 ± 0.15	2.95 ± 0.08
	2	8.81 ± 0.12	6.12 ± 0.14	4.64 ± 0.05	3.94 ± 0.01
	3	9.07 ± 0.19	7.62 ± 0.19	5.84 ± 0.18	4.47 ± 0.06
	4	6.98 ± 0.23	7.37 ± 0.17	5.58 ± 0.12	4.58 ± 0.05

Values are the mean ± s.e.m. l.s.d. ( $\alpha = 0.05$ ) = 0.75 and 0.39 L mg<sup>-1</sup> m<sup>-1</sup> for arable and grassland, respectively.

Leachate SUVA<sub>254</sub> generally decreased ( $P < 0.001$ ) with decreasing  $\theta_g$  in both the arable and grassland soils, indicating that an increasing proportion of non-aromatic organic compounds was released as  $\theta_g$  declined. Such an effect has been shown to occur as a result of changes in the microbial community structure (Schimel et al., 2007) or function (Fierer et al., 2003; Zornoza et al., 2007), leading to the decomposition of group-specific organic compounds and enhanced nutrient release; increased microbial cell lysis (Bottner, 1985; Kieft et al., 1987; Van Gestel et al., 1993a), and the consequent release of lipids, aliphatic moieties of proteins and peptides, and amino sugars (Gregorich et al., 2003; Miltner et al., 2009); decreased structural stability of soil aggregates (e.g. slaking, Adu and Oades, 1978; Denef et al., 2001a), increasing the solubilization of labile organic compounds (Powelson and Jenkinson, 1976); or a combination of these effects in response to soil drying and fast rewetting.

In the arable soil, lower SUVA<sub>254</sub> values occurred in the first PV of all  $\theta_g$  treatments (Table 3.3), indicating a greater proportion of aliphatic compounds. Because the SUVA<sub>254</sub> values of the first PV of the arable soil did not significantly change with soil drying (an increase (l.s.d. < 0.05) occurred only in the 15% treatment),



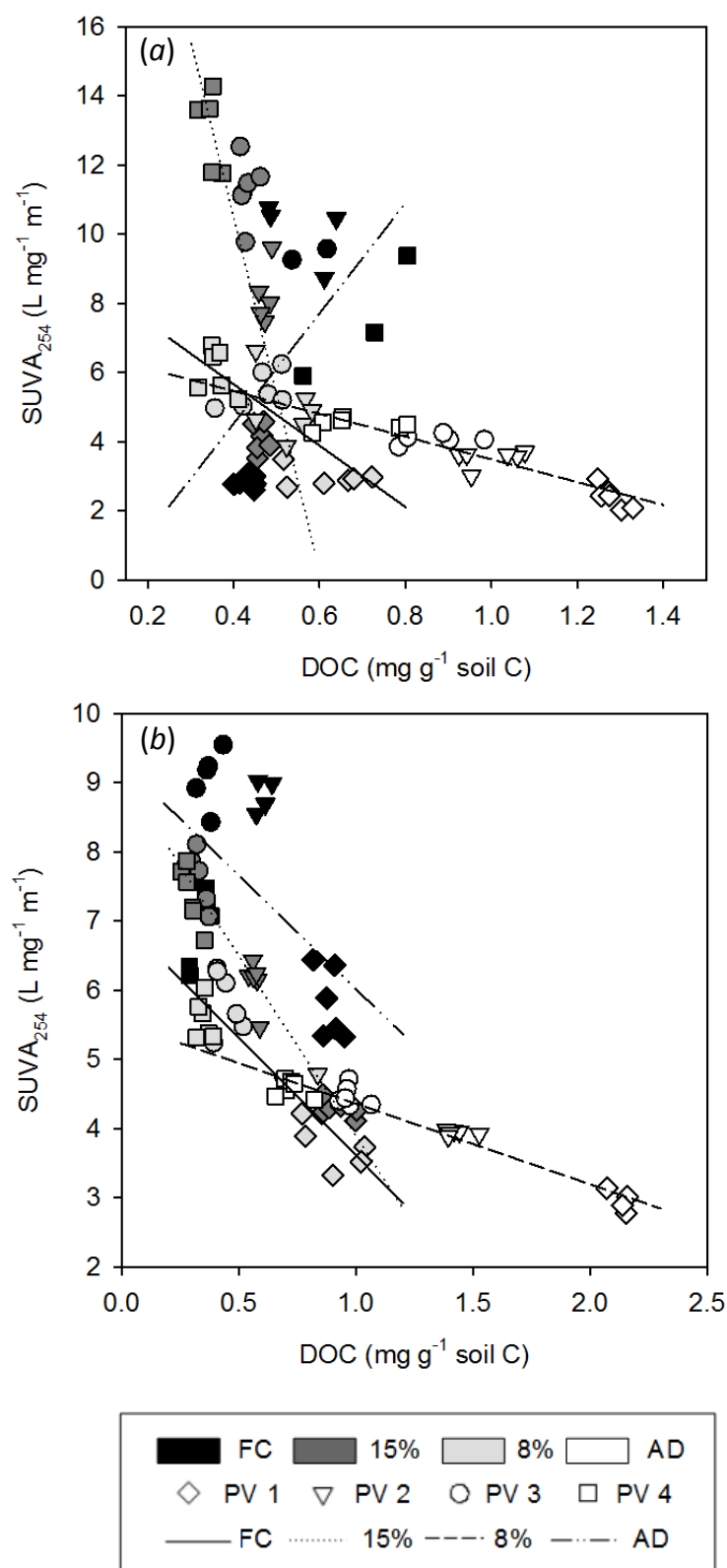
we inferred that factors other than the extent of soil drying had a larger impact on the quality of the OM readily released into solution after rewetting (first PV). For instance, crop residue management and/or accelerated SOM decomposition under cropping conditions might have exposed labile SOM and possibly homogenized the composition of DOM. However, this hypothesis would require further testing to be confirmed.

In the first PV leached from the grassland soil at FC,  $SUVA_{254}$  values indicated a relatively large proportion of aromatic and conjugated olefinic C, but there was a progressive decrease as  $\theta_g$  declined, suggesting that relatively more aliphatic DOM was released with increased soil drying.

Subsequent PVs showed significant changes in the quality of DOM released. The  $SUVA_{254}$  was consistently higher ( $P < 0.001$ ) in the second PV leached, with the magnitude of this increase tending to decline as  $\theta_g$  decreased. Differences in  $SUVA_{254}$  were relatively small or inconsistent in the third and fourth PVs leached, with the exception of the FC treatment, where  $SUVA_{254}$  values were lower in the fourth PV.

In the FC and 15% treatments of the arable soil, the rate of increase in  $SUVA_{254}$ , with respect to changes in DOC concentration, was greater than that in the 8% and AD treatments (Fig. 3.5a). We hypothesize that a net release of aromatic and conjugated olefinic C, rather than a relative decrease of the non-aromatic component, occurred in progressively leached PVs of the FC and 15% treatments compared with the 8% and AD treatments. This hypothesis is supported by the significant increase in absorbance at 254 nm observed after the first PV leached from the FC and 15% treatments of the arable soil (see Fig. A.2a in the Appendix A to this Chapter), as opposed to small or insignificant changes in absorbance at 254 nm that were accompanied by greater decreases in DOC concentrations of progressively leached PVs of the 8% and AD treatments (Fig. 3.5a). In the latter, less marked but significant increases in  $SUVA_{254}$  are likely to reflect an increase of the relative proportion of absorbing moieties, rather than their net release into the percolating solution, resulting from the depletion of the aliphatic DOM fraction released upon drying.

In the grassland soil, the observed increase in  $SUVA_{254}$  spanned large intervals of decreasing DOC concentration (with exception of the fourth PV, Fig. 3.5b). Yet, relatively stable (second PV) and progressively marked decreases (third and fourth PVs) in the absorbance at 254 nm (Appendix A, Fig. A.2b) suggest that, in the grassland soil, increases in  $SUVA_{254}$  were largely driven by the magnitude of DOC decrease with respect to the changes in the relative proportion of absorbing moieties. Interestingly, similar values of absorbance in the FC and AD treatments (e.g. first PV; Fig. A.2b) corresponded to notably different values in  $SUVA_{254}$  (Table 3.3), thus suggesting a dilution of the absorbing component in the AD treatment by the



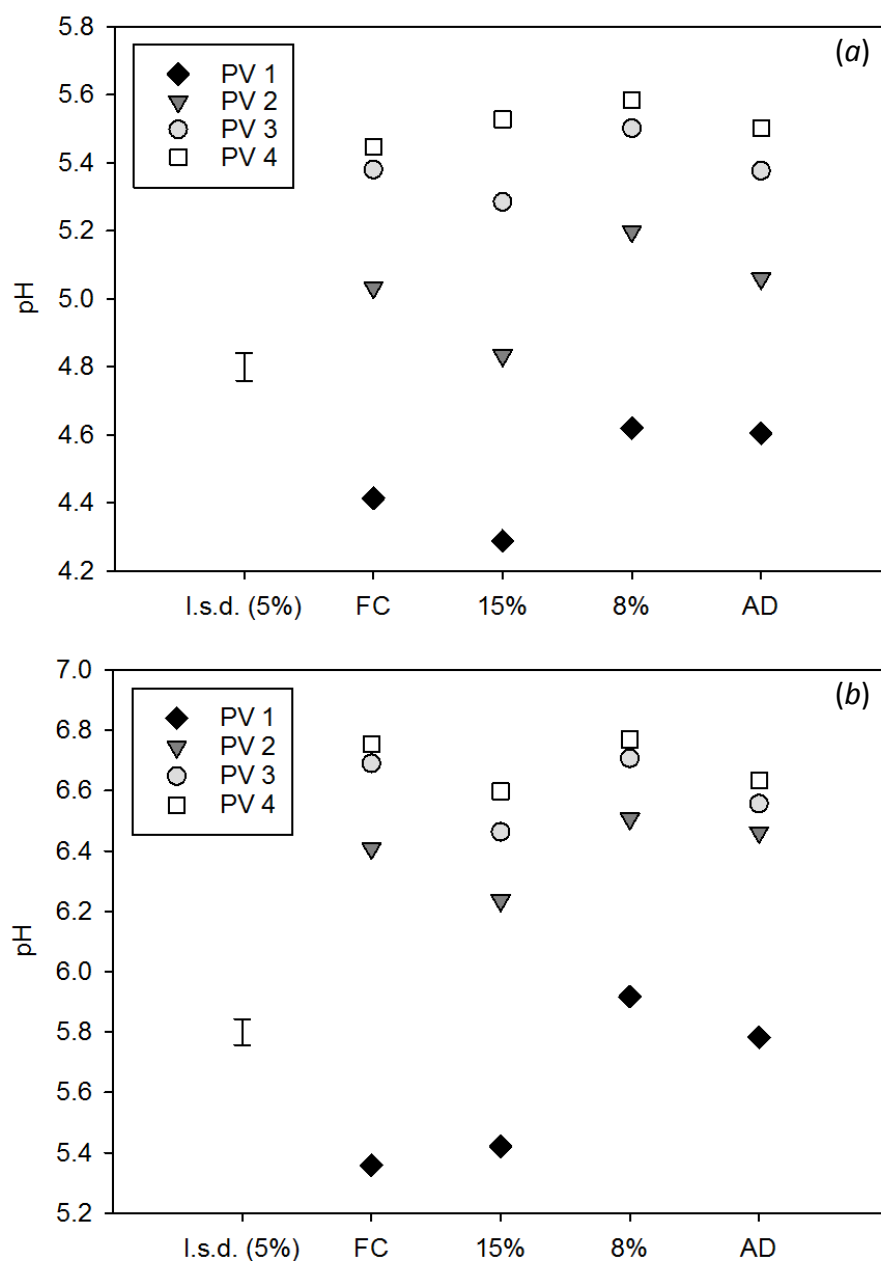
**Figure 3.5:** The relationship between the concentration of dissolved organic C (DOC) and SUVA<sub>254</sub> in leachates of different  $\theta_g$  treatments of an arable (a) and a grassland soil (b). FC, field capacity; AD, air dry; PV, pore volume.

relatively large proportion of non-absorbing aliphatic compounds.

The SUVA<sub>254</sub> values recorded in the arable soil at FC and 15%  $\theta_g$  might be considered exceptionally high (7.5–13 L mg<sup>-1</sup> m<sup>-1</sup>). Although NO<sub>3</sub><sup>-</sup> can interfere with light absorbance at 254 nm at concentrations higher than 40 mg L<sup>-1</sup> (Weishaar et al., 2003), corrections to SUVA<sub>254</sub> were not applied in the current study because potential interferences from large NO<sub>3</sub><sup>-</sup> concentrations (~100 mg L<sup>-1</sup>), detected only in the first PV of the arable soil, would have caused negligible increases (0.01 cm<sup>-1</sup> at 254 nm). Potential interferences due to soluble iron forms (both Fe<sup>2+</sup> and Fe<sup>3+</sup>) were also ruled out, since an increase of 0.4 cm<sup>-1</sup> for 5 mg Fe L<sup>-1</sup> (Weishaar et al., 2003) can be considered negligible for the absorbance values recorded in progressively leached PVs (>2 cm<sup>-1</sup>). Although we do not have available information on leachate Fe concentrations in this study, we assumed that low concentrations of soluble Fe forms occurred (<5 mg L<sup>-1</sup>), as commonly observed for the soil utilized in this study.

The SUVA<sub>254</sub> values reported in most studies pertain to surface waters (in a range of 1 to 6 L mg<sup>-1</sup> m<sup>-1</sup>, Hansen et al., 2016); however, high SUVA<sub>254</sub> values similar to the current study have been reported for soil interstitial waters (>7, Jaffé et al., 2008), litter leachates (>10, Inamdar et al., 2012), and peat soil leachates (~3, Hansen et al., 2016). Furthermore, Rouwane et al. (2018) recently demonstrated a 2-fold increase of SUVA<sub>254</sub> values (up to 10 L mg<sup>-1</sup> m<sup>-1</sup>) in agricultural soils as moderately reducing conditions developed from initially oxic conditions.

The presence and the dynamics of the aromatic moieties, revealed by SUVA<sub>254</sub>, result from differences in soil management influences on OM inputs (quantity and quality) and the ensuing chemical, biological or physical processes that occur as a result. Previous studies reported differences in the molecular composition of SOM under arable or grassland systems, inferring a larger contribution of microbially derived SOM in arable soils, but with greater plant-derived SOM contributions in grassland soils (Guggenberger et al., 1995; Nierop et al., 2001; Martens et al., 2004). In their work on disturbed and undisturbed peatlands, Kalbitz et al. (1999) reported an increased aromaticity of water-soluble fulvic compounds in long-term intensively managed agricultural areas in comparison to extensively used or unused areas. Similarly, Haumaier and Zech (1995) attributed the large presence of aromatic humic acids to the increased mineralization of the litter input due to high microbial activity and cultivation. Arable soils investigated by Nierop et al. (2001) revealed a large presence of heterocyclic N compounds, alkylbenzenes and phenols, which have been linked to strongly humified plant material and microbially altered proteinaceous material (Bracewell and Robertson, 1984). The abundance of such compounds has been linked to greater SOM decomposition and cultivation intensity (Nierop et al., 2001), and possibly poor aggregate protection (Kalbitz et al., 1999). Although we did not further investigate the molecular characteristics of aro-



**Figure 3.6:** The pH of the different pore volumes leached from an arable (a) and a grassland soil (b) at their designated  $\theta_g$  before rewetting. l.s.d. ( $\alpha = 0.05$ ) = 0.08 and 0.09 for arable and grassland, respectively. FC, field capacity; AD, air dry; PV, pore volume.

matic DOM, these prior studies may explain the higher  $\text{SUVA}_{254}$  values observed in leachates of the arable soil. The presence of aromatic moieties, revealed by

SUVA<sub>254</sub>, in the grassland soil of the current study, might have derived from the enzymatic degradation of ligno-cellulose compounds, or have been released with further leaching upon cleavage of stabilized SOM complexes or the breakdown of physically protected macroaggregates (Jongmans et al., 2001), which contain relatively little decomposed OM.

### 3.3.5 Leachate pH

For all treatments and both soils, the leachate pH (Fig. 3.6) was lowest in the first PV (range of 4.3–4.6 in arable and 5.7–5.9 in grassland soil) and highest in the fourth PV leached (5.5–5.6 in arable and 6.6–6.8 in grassland soil). Leachate pH was influenced by an interaction between the  $\theta_g$  treatment and the PV being leached ( $P < 0.001$ ). The FC and 15% treatments tended to have lower pH values than the 8% and AD treatments; however, differences were smaller as leaching progressed. The leachate pH of the first PV leached reflected the soils' initial pH values. Walworth (1992) also reported an increase of  $\sim 0.5$  pH units in soils that were air-dried and rewetted to their original  $\theta_g$  content. Ludwig et al. (1999) explained that, by affecting soil solution composition, mineral dissolution, OM complexation, microbial mineralization, redox reactions and the resulting equilibrium between these processes underlie the air-drying induced increase in pH. Thus, it is possible that changes in DOM quality, as observed from changes in SUVA<sub>254</sub>, may also have contributed to changes in leachate pH.

### 3.3.6 Biodegradability of DOC

Biodegradability of leached DOC (DOC- $C_{min}$ ) varied with  $\theta_g$  treatments in both the arable and grassland soils, with 6.7–20.2% of leached DOC mineralized over 14 days at 20°C (data not shown). Of this, 69–84% was mineralized in the first 7 days of incubation, regardless of land use. There was a general trend for the cumulative DOC mineralized over four PVs to increase with decreasing  $\theta_g$  in the arable soil, with exception of the 8% treatment, where the lowest DOC mineralization was followed by a plateau after 4 days of incubation (data not shown). In the grassland soil, the cumulative DOC mineralized over four PVs decreased in the 15% and 8% treatments, but in the AD treatment was consistent with that measured at FC.

Overall DOC- $C_{min}$  was not affected by the number of PVs leached from the arable soil (Fig. 3.7a). However, there was an effect ( $P < 0.01$ ) of the  $\theta_g$  treatment on the decomposition pattern of DOC of the different PVs: DOC- $C_{min}$  of the fourth PV was lower than that of the first PV leached from the arable soil at both FC and AD, but this trend was not observed for the other  $\theta_g$  treatments. The DOC- $C_{min}$  in the grassland soil (Fig. 3.7b) was generally higher in the fourth than in the first PV leached. In the second and third PVs, there was a trend for DOC- $C_{min}$  to decline

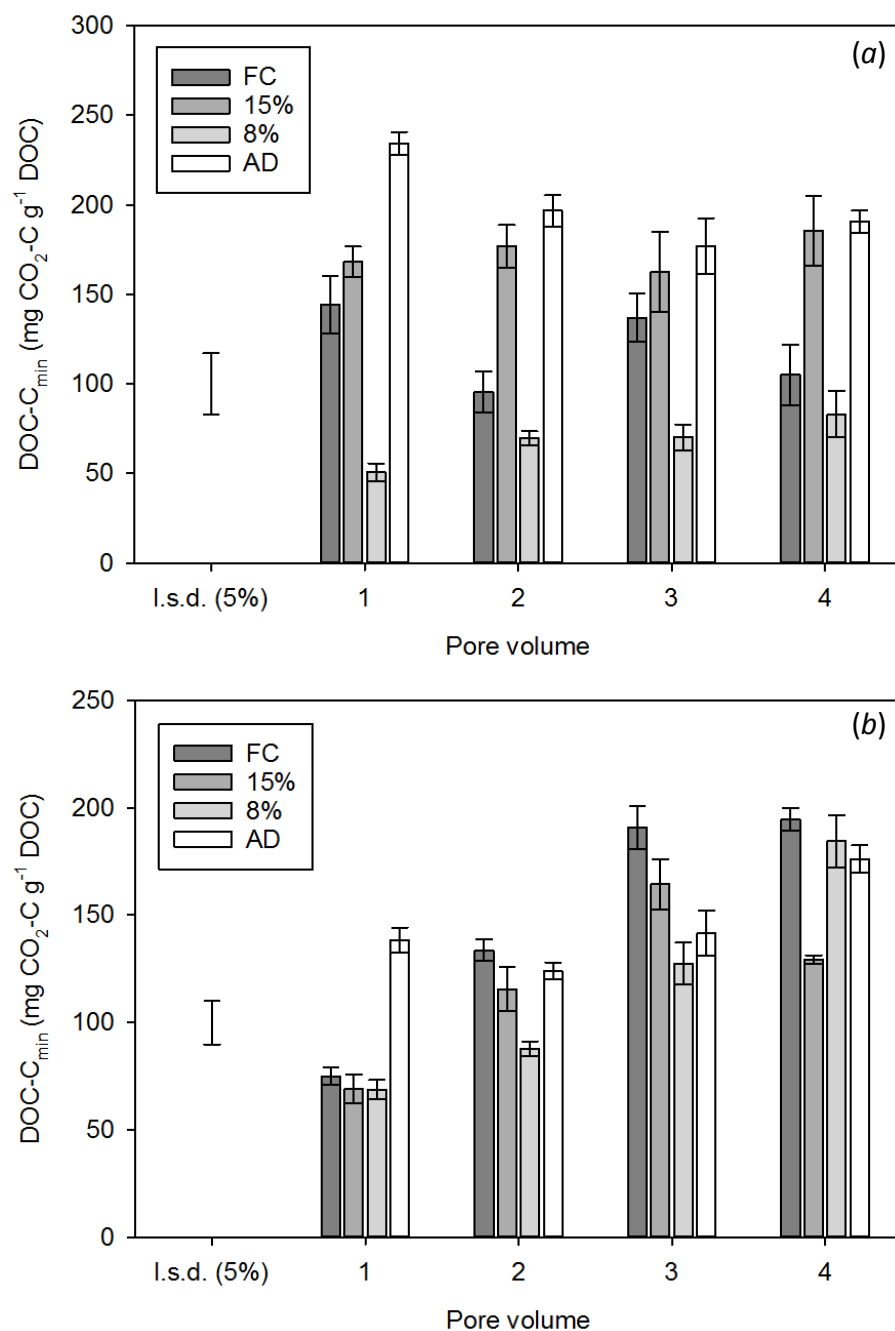
as  $\theta_g$  decreased (15% and 8% treatments), before increasing in the AD treatment. This trend was not as marked or consistent in the first and fourth PV (Fig. 3.7b).

Because arable and grassland soils showed very different patterns of DOC- $C_{min}$ , we further investigated the relationship between DOC- $C_{min}$  and quantitative and qualitative variables. However, no relationships between quantity and quality of DOM and DOC- $C_{min}$  of the arable soil were identified (data not shown); hence, the DOC- $C_{min}$  response was described only for the grassland soil.

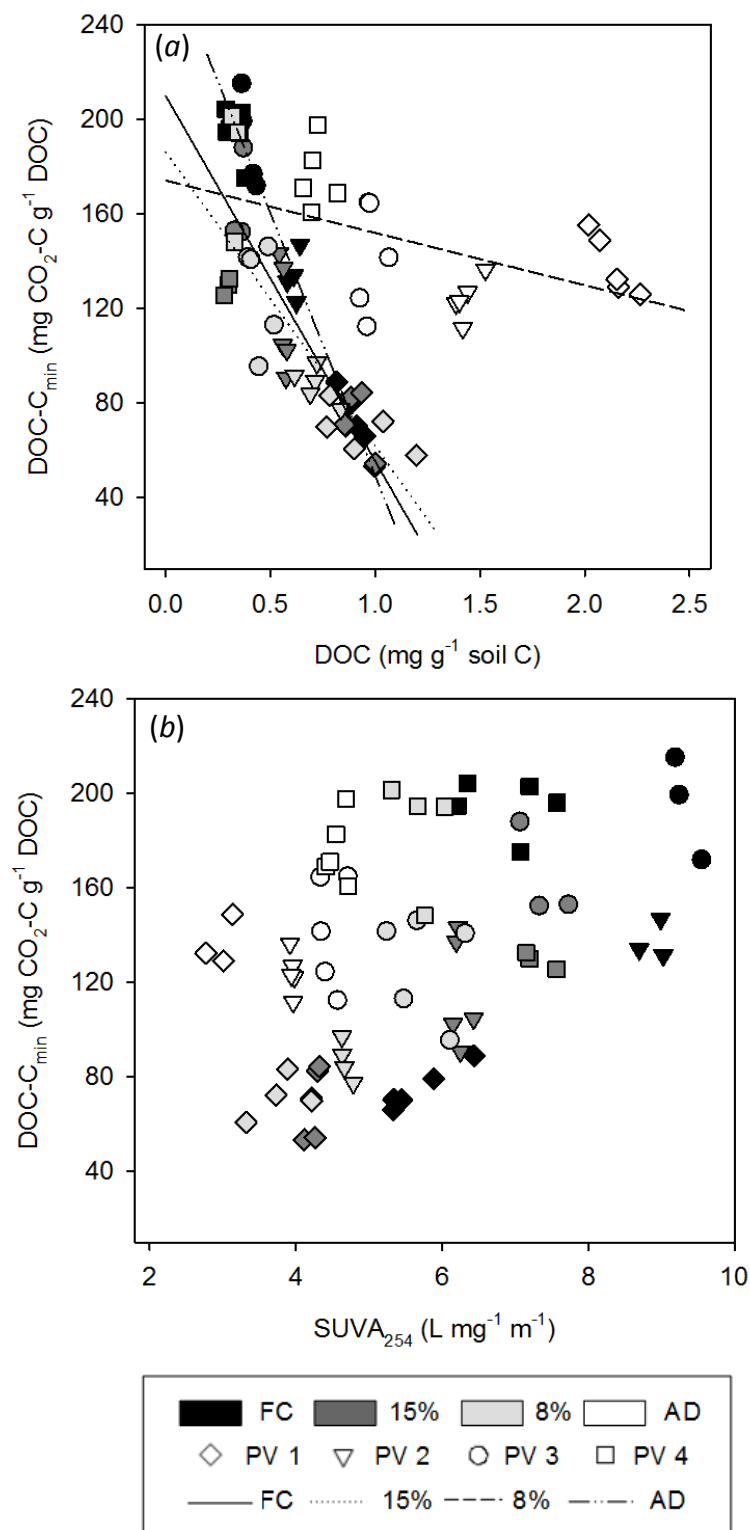
The negative relationship between DOC concentration and DOC- $C_{min}$  in the grassland soil ( $P < 0.001$ , Fig. 3.8a) suggests that the amount of DOC released was not a limiting factor for microbial decomposition, regardless of  $\theta_g$  treatment. Rewetting of the AD soil released a relatively large amount of DOC in the first PV that was less utilized by soil microbiota than the DOC solubilized in the last PV of all  $\theta_g$  treatments, supporting the idea that C availability does not necessarily imply its biodegradability.

We, therefore, investigated whether changes in the quality of DOM, hereby referred to as SUVA<sub>254</sub>, could help to explain the significant differences in biodegradability of sequentially leached DOM. As shown in Fig. 3.8b, soluble organic compounds largely differing in composition and chemical characteristics can be mineralized to the same extent. Similarly, increases in biodegradability of DOC with a comparable SUVA<sub>254</sub> indicate substantial differences in the microbial affinity for these compounds. The response of DOC- $C_{min}$  to changes in SUVA<sub>254</sub> followed a positive association in all  $\theta_g$  treatments ( $P < 0.001$ ), but no consistent model was able to describe this relationship (Fig. 3.8b).

Reasons for the positive relationship between DOC- $C_{min}$  and SUVA<sub>254</sub> require further investigation to identify the molecular composition of leached DOM. Although such data are not available for the current study, we speculate that the (increased) utilization of aromatic compounds was closely related to the (increased) presence of reactive polar functional groups associated with DOM aromatic rings (Korshin et al., 1997). For example, re-oxidation of reduced forms of quinones, either excreted by soil microorganisms (Newman and Kolter, 2000) or derived from the oxidation of lignin (Cory and McKnight, 2005), promotes further lignin breakdown and, thus, the solubility of simplified lignin-derivatives with higher biodegradability that might have been released as leaching progressed. In addition, as suggested by Rouwane et al. (2018), if reducing conditions had developed in the soil cores, as leaching progressed, there may have been sequential release of strongly adsorbed aromatic compounds (Kleber et al., 2007) with a higher oxygenation degree and biochemical reactivity.



**Figure 3.7:** Cumulative mineralization of dissolved organic C ( $\text{DOC-C}_{\min}$ ) in the different pore volumes leached from an arable (a) and a grassland soil (b) at their designated  $\theta_g$  before rewetting. l.s.d. ( $\alpha = 0.05$ ) = 34.08 and 20.31  $\text{mg CO}_2\text{-C g}^{-1} \text{DOC}$  for arable and grassland, respectively. FC, field capacity; AD, air dry; PV, pore volume.



**Figure 3.8:** Cumulative mineralization of dissolved organic C (DOC-C<sub>min</sub>) as a function of DOC concentration (a) and SUVA<sub>254</sub> (b) in leachates of a grassland soil at different  $\theta_g$  before rewetting. FC, field capacity; AD, air dry; PV, pore volume.



## 3.4 Conclusions

The present study showed that the pre-rewetting  $\theta_g$  did not equally affect the quantity and quality, and therefore biodegradability, of DOM leached over sequential PVs from an arable and a grassland soil. In both soils, the DOC release was enhanced to a greater extent only in the AD treatment. Advancing drought levels up to 8%  $\theta_g$  had little if any effect on the bulk of the leached DOC, but revealed a strong effect on the DOC released with sequential leaching. Although the DOC concentrations steadily declined in grassland soil leachates, they changed inconsistently with  $\theta_g$  treatment in the arable soil. Reasons for this may include soil structural changes, microbial interactions or changes in soil redox conditions, affected by an interaction between land use and soil moisture conditions that appears to be evident under leaching conditions, but not in soil extracts. The results, indeed, indicated that the dynamics of a simulated *in situ* DOM release differs from that of a single extraction, commonly performed using standard protocols and artificial soil-to-water ratios. Further work on soil aggregate stability and soil-to-water contact time effects is required to further clarify these results. In addition, repeating such a leaching experiment under high and low temperature conditions might help to evaluate the microbial response to the release of DOC under leaching conditions and infer the microbial interaction.

Sequential soil leaching released an increasing proportion of water-soluble and biodegradable aromatic and/or conjugated olefinic structures. The SUVA<sub>254</sub> values generally decreased with the pre-rewetting  $\theta_g$ , indicating that a larger proportion of aliphatic DOM was released upon rewetting of progressively drier soil conditions. Despite relatively high SUVA<sub>254</sub> values, particularly in the later PVs leached, and differences in DOM quality brought about by soil drying, the biodegradability of DOM over a 14-day period was relatively fast. However, drawing some correlations between the chemical and biological factors influencing DOM production and biodegradability might be achieved only by determining the compositional changes of leached DOM at the molecular level, using more sophisticated analytical methods (e.g. LC-Q-TOF-MS or FT-ICR-MS) and performing an enzymatic characterization. These findings help to further understand the quantitative and qualitative aspects of DOM cycling following soil drying and conditions of prolonged and frequent rainfall or irrigation.

## Supplementary material

The values and  $F$ -probabilities of slope and intercept for the regressions of the individual  $\theta_g$  treatments and the coefficient of determination for each multivariate analysis are presented in the Tables A.1 and A.2 of Appendix A. The figure A.2

in Appendix A shows the UV absorbance at 254 nm for each pore volume leached from an arable ( $a$ ) and a grassland soil ( $b$ ) at four levels of  $\theta_g$ .

## Chapter 4

### **Quantitative and qualitative properties of WEOM fractions in relation to biodegradability: an anion exchange fractionation study**

The first section of this chapter was prepared for submission to the journal  
Soil Biology & Biochemistry and hereby adapted with the approval  
of all co-authors.

Tihana Vujinović, Timothy J. Clough, Denis Curtin, Michael H. Beare,  
Niklas J. Lehto, and Maryam Alavi

The findings presented in the previous Chapter suggested a strong interaction between soil water content and land use on the quantity and quality of organic matter sequentially released in solution. Understanding and predicting the impact of soil moisture on organic matter cycling, thus, requires greater understanding of the fractions responsive to different land uses. The work presented in the current Chapter aims at identifying and characterizing soluble organic matter fractions differing in their polarity, hence reactivity, by means of anion exchange resin treatment.

## 4.1 Biodegradability and qualitative characteristics of WEOM fractions in a soil with different land use histories

### Abstract

There is a need to identify and characterize water extractable organic matter (WEOM) fractions that can elucidate soil organic carbon mineralization responses to land use and management practices. In this study, the uptake of WEOM by anion exchange resin enabled the unretained sub-fraction to be identified, that differed in its mineralization potential when comparing cold (CW) or hot water (HW) fractions of WEOM (referred to as CWEOM and HWEOM, respectively). Extracts were obtained from a soil collected from either a long-term permanent pasture, a no-tillage cropping rotation or a chemical “fallow”, and these were treated with an anion exchange resin to discriminate WEOM according to its charge. The resin treatment removed up to 67% and 86% of the CWEOM and HWEOM, respectively, regardless of land use treatment. Fluorescence fingerprints and specific ultraviolet absorbance at 254 nm revealed the structurally complex and aromatic nature of the negatively charged and/or hydrophobic fraction of WEOM retained by the resin, in contrast to the recently produced, possibly microbially derived, aliphatic constituents of the uncharged/protonated sub-fraction. While the latter was associated with higher biodegradability in the CW extracts, such a relationship was not observed in the HW extracts. Contrary to expectations, the mineralization response was analogous for the CWEOM and HWEOM of whole extracts (non-resin treated) regardless of land use. This may have been due to the large presence of phenolic compounds in the HW extracts, whose mineralization response was primarily led by N-containing compounds. These results suggest the imperative role of plant inputs and substrate quality in regulating microbial outputs and the formation of WEOM sinks and sources.

**Keywords:** Water extractable organic matter, Biodegradation, Anion exchange, Land use, Fluorescence spectroscopy, Random forest

### 4.1.1 Introduction

It is critical to understand the effects of land use on soil organic matter (SOM) dynamics when designing sustainable farm management practices. Ideally, these practices would increase SOM while mitigating carbon (C) losses. Conversion of grasslands into cropping systems has resulted in significant losses of soil organic C

(SOC), the largest terrestrial C pool and reservoir (Lal, 2004). *Vice versa*, the shift from cropping systems back to grasslands has been shown to increase SOC stocks (Guo and Gifford, 2002) resulting from the enhanced SOM content, which aids C sequestration and improves soil stabilization, water storage and infiltration, and soil fertility. Nevertheless, to date, estimates of SOC mineralization, and thus C losses, or sequestration potentials in response to land use change remain inconsistent due to the large array of chemical, physical and biological factors involved (Manzoni and Porporato, 2009).

Numerous attempts have been made to describe functional SOM pools and fractions (McLauchlan and Hobbie, 2004; Haynes, 2005; Leifeld and Kögel-Knabner, 2005; von Lützow et al., 2007) that can explain and predict SOC mineralization responses. For instance, quantitative and conceptual models defining labile and stable SOM pools (based on substrate quality) have been increasingly adopted to estimate the mass of mineralization products and to predict mineralization rate potentials (Parton et al., 2015). However, such models often assume fixed rates of reactivity for only a limited number of pools defined to date (Parton et al., 2015), each of which interacts to a different extent with a variety of other parameters (e.g. temperature) (Ågren, 2000; Kleber et al., 2011). This limits the accuracy of such models when applied to different soil types from different ecosystems (Shirato et al., 2004). More often than not, this rather mechanistic approach requires powerful computational systems (Ågren and Bosatta, 1998), hence limiting model application. Thus, there remains a need for simple and inexpensive methods that are capable of inferring and characterizing actual functional fractions that drive the SOC mineralization response.

Water extractable organic matter (WEOM) has been regarded as the most dynamic and bioactive SOM pool (Bolan et al., 2011). It is, therefore, susceptible to microbial utilization, namely bioavailable, and soil physicochemical alterations in response to land use and management changes (Kalbitz et al., 2000; Chantigny, 2003). However, due to the inherent heterogeneity and the variability of WEOM composition, only a portion of the bioavailable WEOM is readily utilizable and mineralizable within a short timeframe (Cook and Allan, 1992; Marschner and Kalbitz, 2003), hereafter referred to as biodegradable.

Cold (CW) and hot water (HW) extractable SOM fractions are common indicators of bioavailable and biodegradable WEOM (Marschner and Kalbitz, 2003; Haynes, 2005). Although frequently recovered with different extraction procedures (Zsolnay, 2003), the CW fraction serves as a readily available C source for soil microorganisms (Burford and Bremner, 1975). Furthermore, its utilization can also affect soil redox conditions, impacting the release of other gases, such as nitrous oxide and methane. It has been suggested that the bulk of this fraction derives from decomposed SOM, in particular from soil humus (Gregorich et al., 2003). Earlier

studies have shown that changes in soil microclimate (Milorí et al., 2002) and conventional soil management exert a strong impact on the amount of soil “humics” and the degree of SOM humification (Kalbitz et al., 1999; Bayer et al., 2000); it might, therefore, be expected that such changes would be reflected in the size, composition and biodegradability of WEOM contained in the CW fraction (CWEOM).

The HW extractable fraction of SOM has been associated with the readily mineralizable (labile) WEOM (Sparling et al., 1998; Ghani et al., 2003; Gregorich et al., 2003; Curtin et al., 2006; von Lützow et al., 2007), composed of carbohydrates, N-containing compounds, peptides and amides in particular (Leinweber et al., 1995; Ghani et al., 2003; Haynes, 2005), derived from the breakdown of organo-mineral associations, OM hydrolysis and microbial cell lysis (Sparling et al., 1998) boosted by the relatively high extraction temperature. For this reason, it represents the largest fraction of potentially soluble OM, hereafter referred to as HWEOM. Previous studies have shown that the HWEOM responds promptly to changes in land use and management (Ghani et al., 2003). For instance, Haynes and Francis (1993) reported decreasing concentrations of HWEOM in cultivated soils, whereas increases were observed when cropping soils were converted into pasture. In addition, the strong correlation observed between the C content of this fraction and the size, and activity, of the soil microbial biomass pool (Sparling et al., 1998; Ghani et al., 2003) has increasingly promoted its measurement in studies investigating the influence of land use and management on soil quality and its mineralization potential (e.g. Fernández-Romero et al., 2016).

Given that WEOM composition strongly determines its biodegradability (Kalbitz et al., 2003), inconsistencies between WEOM components and biodegradability reported by previous studies mean that the characteristics of WEOM that determine its biodegradability are still not confirmed. Previous studies examining WEOM biodegradability have conducted their investigations on bulk WEOM. It was, therefore, questioned if there could be a sub-fraction more susceptible to mineralization that could define WEOM’s net mineralization rate.

Weak ion exchange (Thurman, 1985) represents a simple method that can separate WEOM fractions based on their charge. WEOM owes its ion exchange properties to polar ionizable functional groups (mainly carboxyl and phenolic, but also carbonyl, alcoholic hydroxyl, methoxyl, and imide groups). The presence of such groups confers upon WEOM a hydrophilic (anionic) and/or amphiphilic character that regulates its bioavailability.

Extensive work has been done on the anion exchange removal efficacy of dissolved organic matter (DOM) from natural waters (Croué et al., 1999; Bolto et al., 2002; Boyer and Singer, 2008), since DOM is a well-known precursor of potentially carcinogenic halogenated organic disinfection by-products (DBPs). Studies that have examined the impact of the anion exchange resin structure on DOM removal

demonstrated that strong base (quaternary ammonium), polyacrylic, macroporous resins, with a lower degree of cross-linking, perform better and are suitable for both low and high molecular weight DOM, given their higher water content, larger exchange capacity and lower selectivity (Fu and Symons, 1990; Bolto et al., 2002). The removal of DOM achieved through the wide range of anion exchange resins examined ranged from 50% to >90% (Fu and Symons, 1990; Afcharian et al., 1997; Bolto et al., 2002). The bulk of this DOM (50–65%) was broadly described as hydrophobic acids, with the remainder comprising of hydrophilic compounds and organic neutrals (Malcolm, 1985; Croué et al., 1999). Further studies have substantiated the correlation between DBPs formation and the aromatic C content of DOM (Edzwald, 1993; Afcharian et al., 1997; Korshin et al., 1997; Singer, 1999). However, qualitative studies have been abandoned due to the low recovery of resin adsorbed DOM, and have been limited to the separation and recovery of specific organic compounds (Kanazawa et al., 2004). Despite the limited understanding of the qualitative aspects of anion exchange resin-fractionated DOM, the ability of anion exchange methods to discriminate DOM/WEOM components according to their charge makes it an interesting and potentially effective tool for studying the nature and biodegradability of soil WEOM.

Laboratory incubations are commonly performed to provide a quantitative estimate of WEOM biodegradability, with functional relevance to SOM dynamics (Schnabel et al., 2002; Gregorich et al., 2003). However, they lack definitive qualitative information about the readily biodegradable WEOM components, whose nature remains controversial. In recent years, UV and fluorescence spectroscopies have been increasingly utilized in DOM and WEOM studies (Chen et al., 2003; Senesi and D'Orazio, 2005; Hudson et al., 2007; Fellman et al., 2010).

UV spectroscopy has been widely adopted to study the chemical and structural properties of DOM/WEOM, and in particular the degree of its aromaticity and structural conjugation (Traina et al., 1990; Novak et al., 1992; Chin et al., 1994). The latter have been strongly correlated with the specific UV absorbance at 254 nm normalized for the DOC concentration ( $SUVA_{254}$ , Weishaar et al., 2003). Hence, comparing  $SUVA_{254}$  values before and after anion exchange resin treatment of WEOM may provide useful information about the character of its fractions.

Several fluorescence indicators have been promoted, with the aim of determining the DOM source (Cory and McKnight, 2005), degree of decomposition (Parlanti et al., 2000; Wilson and Xenopoulos, 2009) and humification (Zsolnay et al., 1999; Ohno, 2002; Milori et al., 2002). Although many of these indicators have emerged from studies on DOM in aquatic environments and humic substances (Miano et al., 1988; Senesi et al., 1991; Stevenson, 1994; Parlanti et al., 2000; McKnight et al., 2001), their utilization in the research of soil DOM and WEOM has gained increasing interest (Kalbitz et al., 1999; Zsolnay et al., 1999; Ohno, 2002; Bu et al., 2010;



Fernández-Romero et al., 2016). In this context, studies investigating the effect of land use and management on the degree of humification (Kalbitz et al., 1999; Milori et al., 2002) have inferred a greater concentration of humified and aromatic components to occur in long-term conventional cropping systems when compared to non-agricultural or low-intensity cropping areas. Yet, correlations between origin and decomposition status of WEOM remained inconclusive, likewise evidence for the relationship between these properties and WEOM's biodegradability potential. Thus, there remain preconceptions about the biodegradability of humified and aromatic moieties, with the idea that WEOM of recent origin and a low degree of decomposition is most biologically active (Wander, 2004), despite increasing evidence arguing against these concepts (Kleber et al., 2011).

To date, conceptual distinctions and correlations between qualitative and quantitative aspects of isolated fractions have been exclusively investigated rather than incorporated into a larger experimental framework. Hence, there remains the need to validate the significance of these qualitative indicators in studies on soil WEOM and to combine the functionality with the structural properties of WEOM fractions.

This study attempted to identify potential predictors of WEOM biodegradability, assessed by standard incubation methods, by means of spectroscopic characterization of anion exchange-fractionated CWEOM and HWEOM. The objectives were to: (i) quantify the effects of land use history on the biodegradability and qualitative characteristics of WEOM, (ii) investigate the potential to use anion exchange resins to identify functionally and biochemically different fractions of WEOM and (iii) determine which aspects of WEOM best explain the variability in mineralizable WEOM.

## 4.1.2 Materials and methods

### Soils

Topsoil samples (0–7.5 cm) were collected in September 2013 from a field trial established to investigate the effects of a long-term ryegrass (*Lolium perenne* L.) and white clover (*Trifolium repens* L.) pasture, a no-tillage crop rotation (barley (*Hordeum vulgare* L.), wheat (*Triticum aestivum* L.), pea (*Pisum sativum* L.) and ryegrass seed crops) and a chemical “fallow” treatment on SOM dynamics. The trial was arranged in a split-plot (9 × 28 m) experimental design with three replicates for each land use treatment. Treatments were established in November 2000 near Lincoln, Canterbury, New Zealand (43° 40' S, 172° 28' E, 5 m above sea level) on a Wakanui silt loam (32% sand, 53% silt, 15% clay, classified as a Mottled Immature Pallic Soil, New Zealand soil classification (Hewitt, 2010); Eutric Siltic Cambisol, IUSS Working Group WRB, 2015). The site had been under irrigated, sheep-grazed, ryegrass/white clover pasture for at least 14 years prior to commencing the trial.

The pasture plots had been maintained as described above pre-trial commencement; the no-tillage plots had been managed under rotational cropping with direct drilling and were ultimately kept under perennial grassland (cv. ‘Grasslands Nui’) after the last cropping season for 1.5 years prior to sampling; the “fallow” plots had been maintained plant-free for the duration of the trial using glyphosate and had not been physically disturbed during the experiment. Further details on the trial, including management information, have been given by Fraser et al. (2013). Eight soil cores ( $\varnothing$  5 cm) were taken along a lengthwise transect in each plot and composited. Samples were then sieved ( $<4$  mm), air-dried and stored for subsequent characterization.

General soil properties are given in Table 4.1. Bulk density was determined following standard methods (Carter and Gregorich, 2008). Total C (TC) and N (TN) were measured on a LECO CNS-2000 elemental analyzer (LECO Corp, St. Joseph, MI, USA). Soil pH was determined in a 1:2.5 soil to deionized (DI) water slurry, vigorously stirred and left to stand for 1 h (Devey et al., 2010). Cation exchange capacity (CEC) was assessed using the ammonium acetate method (Thomas, 1982) and expressed as  $\text{cmol}_c \text{ kg}^{-1}$  soil. For textural analysis, soil samples were dispersed with an ultrasonic vibrator and then separated into sand, silt and clay fractions using a sieving and settling process (Gee and Or, 2002).

### Cold and hot water extractions

Each treatment replicate was extracted in triplicate by adding ultrapure MilliQ water (Millipore,  $18.2 \text{ M}\Omega \text{ cm}^{-1}$ ) and shaking on an orbital shaker at  $20^\circ\text{C}$  (cold water – CW extraction) for 1 h at a soil:water ratio (dry weight basis) of 1:10. The extracts were then centrifuged at  $2831 \text{ g}$  for 5 min and the supernatants were re-centrifuged at  $17\,696 \text{ g}$  for 5 min, to remove any residual particulate material that could hinder subsequent filtration through  $0.45\text{-}\mu\text{m}$  polyvinylidene fluoride (PVDF) membrane filters. The extracted soils were then re-suspended in ultrapure MilliQ water at a soil:water ratio 1:10 and shaken in a water bath at  $80^\circ\text{C}$  (hot water – HW extraction) for 16 h (modified from Gregorich et al., 2003). Hot water extracts were cooled and then centrifuged at  $2831 \text{ g}$  for 5 min, with the supernatants re-centrifuged at  $17\,696 \text{ g}$  for 5 min prior to filtration ( $0.45\text{-}\mu\text{m}$  PVDF membrane filters).

### Anion exchange resin treatment

Five grams of a macroporous, crosslinked polyacrylamide, strong base (Type I), chloride form anion exchange resin (Lanxess Lewatit S 5528) were added to a 100 mL aliquot of either the CW or HW extracts and then shaken (170 rpm) in an orbital shaker either for 2 h (CW) or 61 h (HW). These equilibration times were determined

following pre-testing of the maximum water extractable organic C (WEOC) removal efficacy (see Section B.1 in Appendix B). Resin-treated samples were then filtered through 0.45- $\mu\text{m}$  PVDF syringe filters to separate the extracts from the resin. The fraction remaining after the resin treatment was hereafter referred to as the post-resin treated sub-fraction. The charge density of the WEOM isolates was calculated from the amount of WEOC exchanged for chloride ions released by the resin and expressed in milliequivalents (meq) per gram of WEOC.

### Sample analyses and characterization of WEOM

The WEOC concentrations of the non-resin treated (whole extract) and post-resin treated CW and HW extracts, hereafter referred to as CWEOC and HWEOC, were determined as non-purgeable organic C (NPOC) on a Shimadzu TOC- $V_{CSH}$  analyzer (Shimadzu Corporation, Kyoto, Japan). Total dissolved nitrogen (TDN) was analyzed using the on-line UV-catalyzed persulfate oxidation method (Cabrera and Beare, 1993), while mineral N species ( $\text{NO}_3^-$  and  $\text{NH}_4^+$ ) were quantified colorimetrically on a Lachat QuickChem 8500 Series 2 Flow Injection Analysis System (Lachat Instruments, Loveland, CO, USA). The water extractable organic N concentrations (WEON), hereafter referred to as CWEOC and HWEOC for the CW and HW extracts, respectively, were then calculated by subtracting the mineral N from the TDN.

The quality of WEOM was assessed using a suite of spectroscopic methods. The specific absorption at 254 nm ( $\text{SUVA}_{254}$ , Weishaar et al., 2003) was determined on a UV-Vis SpectraMax 190 microplate reader (Molecular Devices Corp., San Jose, CA, USA). The non-resin treated samples were diluted ten times prior to analysis. In addition to  $\text{SUVA}_{254}$  as a measure of aromaticity, the absorption coefficient at 254 nm ( $a_{254}$ ) was calculated according to Green and Blough (1994). Fluorescence fingerprint indices, namely the fluorescence index (FI, McKnight et al., 2001; Cory and McKnight, 2005), the freshness index ( $\beta:\alpha$ , Parlanti et al., 2000; Wilson and Xenopoulos, 2009) and the Milori humification index ( $A_{465}$ , Milori et al., 2002), were extrapolated from fluorescence spectra collected with a spectrofluorometer (Horiba Aqualog, Horiba Instruments Inc., Kyoto, Japan) over a range of excitation wavelengths from 240 to 600 nm at 3 nm increments and emission wavelengths from 245 to 828 nm (mean slit bandwidth of 4.66 nm). A 10-fold dilution was applied only to the extracts that had not been resin-treated, to set a maximum WEOC working concentration of 35 mg  $\text{L}^{-1}$  and to avoid scattering, inner-filter effects, signal saturation and minimize quenching. Emission intensities were corrected *a posteriori* for the dilution factor applied. The FI, previously correlated with the source of DOM (microbial or plant and SOM-derived), was calculated according to Cory and McKnight (2005) and adapted to this experiment as the ratio of the emission intensities recorded at 471 and 522 nm, when excited at 369 nm. The  $\beta:\alpha$

index was utilized as an indicator of recently derived DOM ( $\beta$ ) as opposed to older and highly decomposed DOM ( $\alpha$ ), and was calculated as the emission intensity at 379 nm divided by the maximum emission intensity recorded between 420 and 434 nm, at 309 nm excitation. Since the inspection of the fluorescence spectra did not show any peak in the region of emission (435–480/300–345 nm) defining the humification index at 255 nm excitation (HIX, *sensu* Zsolnay et al., 1999; modified by Ohno, 2002), the  $A_{465}$ , calculated as the intensity area under the emission spectra at 465 nm excitation, was chosen as being the appropriate humification index for this study. The integrated values, expressed in Raman units (R.U.), were then normalized for the WEOC concentration ( $\text{mg L}^{-1}$ ) in the sample.

Total dissolved phenols in CW and HW extracts were quantified with the Folin-Ciocalteu's reagent (Ohno and First, 1998) and expressed as salicylic acid equivalents per unit of WEOC.

The pH of the extracts was measured directly with a calibrated Thermo Scientific™ Eutech PC 450 pH-meter. Chloride, fluoride, sulphate and phosphate were measured using an ion chromatography system (Dionex DX-2100) fitted with an IonPac AS18 ( $4 \times 250$  mm) analytical column and an IonPac AS18 ( $50 \times 4$  mm) guard column at 30°C and a flow rate of  $1 \text{ mL min}^{-1}$ . A standard stock solution with mixed anions ( $1000 \text{ mg L}^{-1}$ ) was used to prepare a calibration curve over the range of 0.02–100  $\text{mg L}^{-1}$ . For this analysis, the analytical replicates (triplicates) of all extracts were bulked together to keep the costs within the available financial resources.

All extracts were analyzed fresh and stored at 4°C overnight prior to subsequent incubation.

### Biodegradation assay

To account for the fact that biodegradable WEOC comprises of mineralizable and microbially assimilable WEOM, both  $\text{CO}_2$  production and WEOM disappearance (McDowell et al., 2006) were accounted for. For the purposes of this study, the  $\text{CO}_2$  production was referred to as WEOC mineralization ( $\text{WEOC}_{\text{min}}$ ), whereas the WEOM disappearance as WEOM utilization.

Biodegradable WEOC was assessed by incubating 40 mL of either the CW or HW extract, or the respective post-resin treated extract, with 2 mL of N-free nutrient solution and 100  $\mu\text{L}$  of fresh soil inoculum ( $1.77 \text{ mg C g}^{-1}$  soil) in sealed, acid-washed, 100 mL Schott bottles at 25°C for 14 days. To prepare the inoculum, 4 g of agricultural soil (Udic Haplustept, USDA 1999) were shaken with 45 mL of deionized water and incubated at 35°C for 24 h (Gregorich et al., 2003). The inoculum was then re-suspended and the upper 30 mL were transferred into a clean 50 mL Falcon™ tube after 1 h. Batches of CW and HW extracts, along with the respective post-resin treated extracts, were incubated within 24 h of extraction.

Forty milliliters of either MilliQ water, a glucose solution (20 mg C L<sup>-1</sup> for the CWE and 80 mg C L<sup>-1</sup> for the HWE) or a phenol solution (12.2 mg C L<sup>-1</sup>) were also added with 2 mL of N-free nutrient solution and 100  $\mu$ L of fresh soil inoculum and run in triplicate as blanks, positive and negative controls for each batch incubation. The phenol solution comprised: quercetin (2.1 mg C L<sup>-1</sup>), coumaric (2.8 mg C L<sup>-1</sup>), syringic (1.6 mg C L<sup>-1</sup>), cinnamic (2.5 mg C L<sup>-1</sup>), nicotinic (0.2 mg C L<sup>-1</sup>) and salicylic (3 mg C L<sup>-1</sup>) acids.

Values of WEOC<sub>min</sub> were determined as cumulative CO<sub>2</sub> produced over a period of 14 days and expressed as CO<sub>2</sub>-C produced per unit of WEOC present in the extract at the start of the incubation (Baldock and Broos, 2012). Headspace gas samples were withdrawn 3, 8 and 14 days after the beginning of the incubation using a 20 mL gas-tight syringe. These were injected into a LI-COR Li-7000 infrared gas analyzer (IRGA) to determine CO<sub>2</sub> concentrations. Before being returned to the incubator, the incubation bottles were flushed to re-establish an atmospheric CO<sub>2</sub> concentration. At the end of the incubation, samples were filtered (0.45- $\mu$ m PVDF syringe filters) and analyzed for NPOC, TDN and mineral N, as described above, to quantify, respectively, the amount of WEOC and WEON utilized. In addition, the ratio of WEOC respired as CO<sub>2</sub> to the WEOC utilized was calculated as a measure of metabolic efficiency and expressed as cumulative CO<sub>2</sub>-C produced per unit of WEOC utilized.

To determine dissolved CO<sub>2</sub>, 10 mL of each post-incubation sample were transferred to 50 mL Schott bottles, whereupon 2 mL of 2 M HCl were added and the evolved CO<sub>2</sub> analyzed on the IRGA after 2 h. Dissolved CO<sub>2</sub> was then calculated by applying the Henry's law of ideal gas, and the concentration of bicarbonate species derived from the dissociation equilibria equation at a given pH (Dahlgren et al., 1997).

### Statistical analyses

A Linear Mixed Model (LMM) was fitted with Restricted Maximum Likelihood (REML) estimates on the quantitative and quantitative variables (WEOC (mg g<sup>-1</sup> soil C), WEON (mg g<sup>-1</sup> soil N), SUVA<sub>254</sub> (L mg<sup>-1</sup> m<sup>-1</sup>), a<sub>254</sub> (m<sup>-1</sup>), phenols (meq salicylic ac. mg<sup>-1</sup> WEOC), FI,  $\beta:\alpha$ , A<sub>465</sub> (L mg<sup>-1</sup> R.U.) and pH), on the resulting WEOC mineralized (WEOC<sub>min</sub> as cumulative mg CO<sub>2</sub>-C g<sup>-1</sup> WEOC), as well as on WEOM utilized ( $\Delta$ WEOC and  $\Delta$ WEON) and metabolic efficiency (cumulative CO<sub>2</sub>-C produced per unit of WEOC utilized). Fixed effects included land use, extracted fraction, resin treatment and their mutual and 3-way interactions, while the random effects considered plot with land use and subsample replicate as a nested effect. The repeated structure of the subsamples per plot and land use was considered in the model. The response variates were log transformed to satisfy the model's assumptions. The mean and confidence limits were then back-transformed

and reported.

General Analysis of Variance (ANOVA) was used to assess the land use effect on the bulk density, TC, TN, pH, CEC, and absolute amounts of WEOC and WEON released ( $\text{mg g}^{-1}$  soil). The response variates were checked to meet ANOVA assumptions. Variability within the treatments was expressed by the standard error of difference of means (s.e.d.). Statistical significance was set at the 5% level and expressed by the least significant differences (l.s.d.).

Random Forest (RF) with regression trees was used for dimensionality reduction and to infer the importance of the quantitative and qualitative variables on  $\text{WEOC}_{\min}$ . Training and validation (10-fold) were performed by fitting the model on two blocks and validating the model on the third block on each round. The optimal model was selected using the smallest root mean square error (RMSE) and finally, all nine quantitative and qualitative variables were included. The importance of the selected variables was expressed by the Increased Mean Square Error (%IncMSE), measuring the increase in the estimated generalization error when values for a specific variable are randomly permuted without altering the other variables. Also, the Increased Node Purity Index (IncNodePurity) was used, measuring the mean reduction in the residual sum of squares (increase in node purity or homogeneity) due to splits by a given variable over all trees.

The REML LMM fitting and ANOVA were performed using the Genstat statistical package version 18.0. The RF modeling was run in the R environment (R Core Team, 2018). Graphs were prepared using R version 3.5.1.

### 4.1.3 Results

#### Soil characteristics

Soil properties varied with land use and spatial variability (Table 4.1). Soil bulk density was lower in the pasture and no-tillage soils than in the fallow soil. The TC and TN were highest in the pasture soil and lowest in the fallow soil, while intermediate values were measured in the no-tillage soil. The same pattern was observed for the pH, with  $\sim 1$  pH unit difference between the pasture and fallow soils. The CEC was higher in the pasture soil than the fallow, with no difference between the fallow and no-tillage soils.

#### Cold and hot water extractable organic carbon

The amount of both CWEOC and HWEOC was strongly influenced by land use, with greater concentrations in the pasture soil than in the no-tillage and fallow soils (Table 4.2). The WEOC declined by  $\sim 29$  and  $56\%$  (CWEOC) and  $\sim 20$  and  $50\%$

**Table 4.1:** General properties of surface soils (0–7.5 cm) kept under a pasture, a no-tillage or a fallow regime for 13 years.

Land use	Bulk density (g cm <sup>-3</sup> )	TC (g kg <sup>-1</sup> )	TN (g kg <sup>-1</sup> )	pH (H <sub>2</sub> O)	CEC, pH 7 (cmol <sub>c</sub> kg <sup>-1</sup> soil)
Pasture	1.2	35.3	3.2	6.0	16
No-tillage	1.2	28.6	2.6	5.3	14
Fallow	1.4	20.8	1.9	5.0	13
s.e.d.	0.03	1.25	0.11	0.05	1.05
l.s.d.	0.07	3.06	0.27	0.13	2.58

Total carbon (TC); Total nitrogen (TN); Cation exchange capacity (CEC). Values are means ( $n = 3$ ) of field replicates; s.e.d = standard error of differences of means; l.s.d. = least significant difference ( $\alpha = 0.05$ ).

(HWEOC) in the no-tillage and fallow soils, respectively, when compared to the pasture soil.

For all non-resin treated extracts, the release of WEOC per unit of soil C was also largest in pasture and least in fallow soil (Table 4.3), but differences were only statistically significant for the CW extracts, as a result of a 3-way interaction between land use, extracted fraction and resin treatment (Table 4.4).

The resin treatment removed 64–67% of the CWEOC and 80–86% of the HWEOC (Table 4.3). There was a strong linear relationship between the amount of WEOC retained by the resin and chloride released (Appendix B, Fig. B.1). However, the slope of this relationship differed for CW and HW extracts (0.7 and 0.48 for CW and HW extracts, respectively). Based on the quantity of chloride released, the (mean) charge density of the resin-retained WEOC was estimated at  $\sim 25$  meq g<sup>-1</sup> WEOC for the CW extracts and  $\sim 11$  meq g<sup>-1</sup> WEOC for the HW extracts. The concentrations of other anions in solution were relatively low (Appendix B, Tables B.1a and B.1b), with the highest values measured in the HW extracts (fluoride  $< 0.077$  meq L<sup>-1</sup>, sulfate  $< 0.065$  meq L<sup>-1</sup>, phosphate  $< 0.2$  meq L<sup>-1</sup>).

As a result of an interaction between land use, extracted fraction and resin treatment (Table 4.4), the WEOC in the post-resin treated sub-fraction also reflected changes induced by the land use (Table 4.3); however, the WEOC was lower only in the fallow soil.

The HWEOC concentrations were 3 to 4-fold higher than the CWEOC concentrations ( $P < 0.001$ , Table 4.3). The same pattern was observed following resin

treatment, the HWEOC being 1 to 2.5-fold higher than the CWEOC. The HWEOC represented 77–79% of the total WEOC extracted (CW + HW), while in the post-resin treated sub-fraction it made up 62–67% of the total WEOC recovered.

**Table 4.2:** Absolute amounts of WEOC and WEON released following cold (CW) and hot water (HW) extractions of soils from different land use treatments.

Land use	WEOC (mg kg <sup>-1</sup> soil)		WEON (mg kg <sup>-1</sup> soil)	
	CW	HW	CW	HW
Pasture	532	1785	38	175
No-tillage	378	1427	21	118
Fallow	236	887	14	66
s.e.d.	25.2	87.6	1.9	9.1
l.s.d.	70	243	5.3	25.3

Values are means ( $n = 9$ ); s.e.d = standard error of differences of means; l.s.d. = least significant difference ( $\alpha = 0.05$ ).

Across land uses, there was a high correlation between the CWEOC and HWEOC ( $r = 0.82$ ,  $P < 0.001$ ). However, while still significant, this correlation was not as strong in the post-resin treated sub-fraction ( $r = 0.54$ ,  $P = 0.004$ ).

### Cold and hot water extractable organic nitrogen

The quantity of WEON varied among land uses, the pasture soil having greater values than the no-tillage and fallow soils (Table 4.2). The WEON declined by ~44 and 65% (CWEON) and ~33 and 65% (HWEON) in the no-tillage and fallow soils, respectively.

The WEON extracted per unit of soil N also differed with land use (Table 4.3): both CWEON and HWEON were highest in the pasture and lowest in the fallow soil treatments, reflecting the same trend observed for WEOC.

The resin treatment removed 58–67% and 83–87% of the initial CWEON and HWEON concentrations, respectively (Table 4.3). In all the post-resin treated sub-fractions, WEON was highest in the pasture and lowest in the fallow soil, but differed significantly only in the fallow soil, due to a mutual interaction between land use, extracted fraction and resin treatment (Table 4.4).



**Table 4.3:** Quantitative and qualitative properties of non- and post-resin treated cold (CW) and hot water (HW) extracts, as affected by the land use. Values are means ( $n = 9$ ) within estimated confidence intervals (in brackets). ND = not detected.

	Resin treatment	Pasture			No-tillage			Fallow	
		CW	HW		CW	HW		CW	HW
WEOC (mg g <sup>-1</sup> soil C)	non-	15.1 (13.5, 16.7)	50.5 (45.4, 56.1)		13.2 (11.9, 14.7)	49.8 (44.8, 55.3)		11.3 (10.2, 12.6)	42.5 (38.3, 47.3)
	post-	4.9 (4.4, 5.5)	10.0 (9.0, 11.1)		4.8 (4.3, 5.4)	8.3 (7.5, 9.2)		3.8 (3.4, 4.2)	6.1 (5.5, 6.8)
WEON (mg g <sup>-1</sup> soil N)	non-	11.9 (10.6, 13.4)	54.5 (48.4, 61.4)		8.4 (7.4, 9.4)	46.2 (41.1, 52.0)		7.0 (6.2, 7.9)	33.9 (30.1, 38.1)
	post-	3.9 (3.5, 4.4)	8.9 (7.9, 10.0)		3.5 (3.2, 4.0)	5.8 (5.1, 6.5)		2.8 (2.5, 3.1)	4.5 (4.0, 5.1)
SUVA <sub>254</sub> (L mg <sup>-1</sup> m <sup>-1</sup> )	non-	2.8 (2.5, 3.1)	2.3 (2.0, 2.6)		2.3 (2.1, 2.6)	1.9 (1.7, 2.1)		2.8 (2.5, 3.1)	2.0 (1.8, 2.2)
	post-	0.6 (0.5, 0.6)	0.7 (0.7, 0.8)		0.5 (0.5, 0.6)	0.6 (0.6, 0.7)		0.8 (0.7, 0.9)	0.6 (0.6, 0.7)
a <sub>254</sub> (m <sup>-1</sup> )	non-	332.7 (293.5, 377.2)	912.7 (805.1, 1034.7)		199.8 (176.2, 226.4)	598.3 (527.8, 678.3)		147.1 (129.8, 166.8)	397.9 (351, 451.1)
	post-	21.9 (19.4, 24.9)	58.5 (51.6, 66.3)		16.3 (14.4, 18.5)	32.9 (29.1, 37.3)		13.5 (11.9, 15.3)	18.2 (16, 20.6)
Phenols (meq Salicylic ac. mg <sup>-1</sup> WEOC)	non-	0.23 (0.20, 0.27)	0.56 (0.48, 0.66)		0.21 (0.18, 0.25)	0.55 (0.47, 0.65)		0.14 (0.12, 0.16)	0.31 (0.26, 0.37)
	post-	ND	ND		ND	ND		ND	ND
FI	non-	1.31 (1.29, 1.34)	1.42 (1.40, 1.45)		1.36 (1.34, 1.39)	1.51 (1.48, 1.55)		1.36 (1.33, 1.39)	1.50 (1.46, 1.53)
	post-	2.09 (2.05, 2.13)	1.82 (1.79, 1.86)		2.09 (2.04, 2.13)	1.85 (1.82, 1.89)		2.01 (1.97, 2.05)	1.94 (1.90, 1.98)
β:α	non-	0.46 (0.45, 0.48)	0.48 (0.46, 0.50)		0.50 (0.48, 0.52)	0.52 (0.51, 0.54)		0.49 (0.47, 0.51)	0.50 (0.49, 0.52)
	post-	0.93 (0.90, 0.96)	0.84 (0.81, 0.87)		0.97 (0.94, 1.01)	0.92 (0.89, 0.95)		0.98 (0.95, 1.02)	0.99 (0.96, 1.03)
A <sub>465</sub>	non-	0.53 (0.49, 0.57)	0.29 (0.27, 0.31)		0.48 (0.44, 0.51)	0.23 (0.22, 0.25)		0.69 (0.64, 0.74)	0.35 (0.33, 0.38)
	post-	0.02 (0.02, 0.02)	0.06 (0.06, 0.06)		0.03 (0.02, 0.03)	0.06 (0.05, 0.06)		0.04 (0.03, 0.04)	0.06 (0.06, 0.06)
pH	non-	6.0 (5.8, 6.2)	5.9 (5.8, 6.1)		5.5 (5.3, 5.6)	4.9 (4.8, 5.1)		5.5 (5.3, 5.7)	5.0 (4.9, 5.2)
	post-	5.0 (4.8, 5.1)	5.4 (5.2, 5.5)		4.8 (4.6, 4.9)	4.6 (4.5, 4.8)		4.9 (4.8, 5.1)	4.5 (4.3, 4.6)

**Table 4.4:** REML output indicating the statistical significance of land use, extraction method, resin treatment and their interaction on the quantitative and qualitative variables investigated.

	WEOC (mg g <sup>-1</sup> soil C)	WEON (mg g <sup>-1</sup> soil N)	SUV <sub>A254</sub> (L mg <sup>-1</sup> m <sup>-1</sup> )	a <sub>254</sub> (m <sup>-1</sup> )	Phenols (meq Salicylic ac. mg <sup>-1</sup> WEOC)	FI	β:α	A <sub>465</sub>	pH
Land use	0.009	0.001	0.099	<0.001	0.002	0.078	0.005	<0.001	<0.001
Extraction	<0.001	<0.001	<0.001	<0.001	<0.001	0.657	0.349	0.005	<0.001
Treatment	<0.001	<0.001	<0.001	<0.001		<0.001	<0.001	<0.001	<0.001
Land use x Extraction	0.466	0.037	<0.001	<0.001	0.637	<0.001	0.041	<0.001	<0.001
Land use x Treatment	0.004	0.987	0.017	0.848		0.001	0.003	<0.001	0.025
Extraction x Treatment	<0.001	<0.001	<0.001	<0.001		<0.001	<0.001	<0.001	0.001
Land use x Extraction x Treatment	<0.001	<0.001	0.003	<0.001		<0.001	0.008	<0.001	0.017

Across all fractions and land uses, WEON represented the largest proportion of the TDN (data not shown). The proportion of CWEON varied from 56 to 76%, being lowest in the fallow and highest in the no-tillage soil, whereas the HWEON made up 86% of the TDN, regardless of land use. In contrast, WEON comprised 80% (pasture) to more than 90% (fallow and no-tillage) of the TDN in the CW post-resin treated sub-fraction, while it ranged from 39% (pasture) to 53% (fallow) of the TDN in the HW post-resin treated sub-fraction. While  $\text{NO}_3^-$  comprised the remaining TDN of the CW extracts, and decreased below the detection limit following the resin treatment,  $\text{NH}_4^+$  remained the predominant mineral N form in both the non-treated and post-resin treated HW extracts.

Regardless of land use and resin treatment, the HWEON was significantly larger than the CWEON ( $\sim 5$  to 6-fold in the non-resin treated and 2-fold in the resin-treated extracts, Table 4.3). The HWEON comprised 82–85% of the total WEON extracted (CW + HW) and represented 62–70% of the total WEON in the post-resin treated sub-fraction. The CWEON and HWEON correlated strongly ( $r = 0.91$ ,  $P < 0.001$ ), and this correlation was still moderately reflected in the post-resin treated sub-fraction ( $r = 0.60$ ,  $P = 0.001$ ).

### Qualitative characteristics

There was no difference in  $\text{SUVA}_{254}$  values due to land use ( $P = 0.099$ ). However, there was a tendency for  $\text{SUVA}_{254}$  of both CW and HW extracts to be highest in the pasture soil and lowest in the no-tillage soil (Table 4.3), but to vary inconsistently in the fallow soil, due to an interaction between land use and extracted fraction (Table 4.4). The  $\text{SUVA}_{254}$  values for CW extracts were higher than those for HW extracts ( $P < 0.001$ , Table 4.3).

The resin treatment strongly reduced  $\text{SUVA}_{254}$  in both the CW and HW extracts ( $P < 0.001$ , Table 4.3) to values  $\leq 0.8 \text{ L mg}^{-1} \text{ m}^{-1}$ . Although differences in the  $\text{SUVA}_{254}$  values of the post-resin treated sub-fraction were subtle, and overall statistically not significant, there was an interaction between the land use and the resin treatment altering the  $\text{SUVA}_{254}$  trends observed for the post-resin treated sub-fraction with respect to those in the non-resin treated extracts ( $P = 0.02$ , Table 4.3). Furthermore,  $\text{SUVA}_{254}$  responses to the resin treatment were also a function of the extracted CW or HW fraction, as well as of the 3-way interaction between land use, extracted fraction and resin treatment (Table 4.4). As such, there was a tendency for  $\text{SUVA}_{254}$  to be highest in the CW post-resin treated sub-fraction of the fallow soil and, in contrast, in the HW post-resin treated sub-fraction of the pasture soil.

The absorption coefficient at 254 nm ( $a_{254}$ ) was affected by land use, extracted fraction, resin treatment and their interaction (Table 4.4). The  $a_{254}$  values were higher in the HW extracts (171–200% relative to the CW extracts) and highest

in pasture soil and lowest in fallow soil (Table 4.3). Values for  $a_{254}$  decreased significantly following resin treatment, but again reflected the trends due to land use and extracted fraction (Table 4.3).

The total phenolic concentration was higher in the HW extracts than in the CW extracts ( $P < 0.001$ , Table 4.3). Phenolic-C comprised  $\sim 31\text{--}56\%$  and  $\sim 14\text{--}23\%$  of the HWEOC and CWEOC, respectively. For both extractions, total phenols were lowest in the fallow soil ( $P = 0.002$ ), but did not differ between the pasture and no-tillage soils. Total phenolic concentrations in the post-resin treated samples were below the detection limit.

The fluorescence index (FI) was generally not affected by land use ( $P = 0.08$ , Table 4.3) and did not differ, statistically, between CW and HW extracts ( $P = 0.66$ ). However, a 3-way interaction between land use, extracted fraction and resin treatment (Table 4.4) showed that values for FI were greater in the non-resin treated HW extracts, with respect to the CW extracts, and were, *vice versa*, lower in the post-resin treated HW extracts than in the post-resin treated CW extracts. Similarly, an interaction occurred between land use and extracted fraction (Table 4.4) such that FI was lower in the HW extracts of the pasture soil than in the no-tillage or fallow soils, but was not affected by land use in the CW extracts. For both CW and HW extracts, FI was much higher in the post-resin treated sub-fraction than in the non-resin treated extracts ( $P < 0.001$ , Table 4.3) and generally greater in CW than in HW post-resin treated extracts. The largest difference between the CW and HW post-resin treated sub-fractions ( $\sim 0.3$  FI units) was observed in the pasture soil, with the smallest ( $< 0.1$  FI units) in the fallow soil.

The fluorescence derived freshness index ( $\beta:\alpha$ ) increased strongly following the resin treatment and did not differ between CW and HW extracts (Table 4.3). Although there was statistical evidence of a land use effect on values for  $\beta:\alpha$  (Table 4.4), which had a tendency to be lower in the pasture soil when compared to the no-tillage and fallow soils, the interaction between land use, extracted fraction and resin treatment showed that this effect was mainly driven by the tendency of  $\beta:\alpha$  to be lower in the post-resin treated sub-fraction of the pasture soil, but to remain consistent across land uses in the non-resin treated fractions (Table 4.3).

For all soils and extracts, values for  $A_{465}$  decreased greatly after the resin treatment (Table 4.3). In the non-resin treated extracts, the  $A_{465}$  was generally lower in the HW extracts ( $P < 0.001$ ) and higher in the fallow soil ( $P < 0.001$ ). Similar trends were observed in the post-resin treated extracts, although, even if statistically significant, differences were relatively small.

## pH

As with the soil pH (Table 4.1), values of pH measured in the CW and HW extracts (Table 4.3) were higher for pasture soil than for the no-tillage and fallow soils ( $\sim 0.5$

to 1.0 pH unit higher for CW and HW, respectively). The pH values of the CW extracts from the fallow and no-tillage soils were  $\sim 0.5$  units higher than in the corresponding HW extracts, while in the pasture soil, the pH did not vary between the extracted fractions.

The pH decreased following the resin treatment, and this decrease was larger in the CW than in the HW post-resin treated sub-fractions (0.6 (fallow) to 1.0 (pasture) pH unit in the CW extracts and 0.3 (no-tillage) to 0.5 (pasture and fallow) pH unit in the HW extracts). However, while the land use did not affect the pH of the CW post-resin treated sub-fraction, the HW post-resin treated sub-fraction of the pasture soil had higher pH values when compared to the fallow and no-tillage soils.

### WEOM biodegradability

Between 32–34% of the CWEOC and HWEOC was mineralized after 14 days (Fig. 4.1). In contrast, only 3–5% of the glucose control (for the glucose at 20 mg L<sup>-1</sup> and 80 mg L<sup>-1</sup>, respectively) and  $\sim 1\%$  of the phenolic control was mineralized within the same time frame (data not presented).

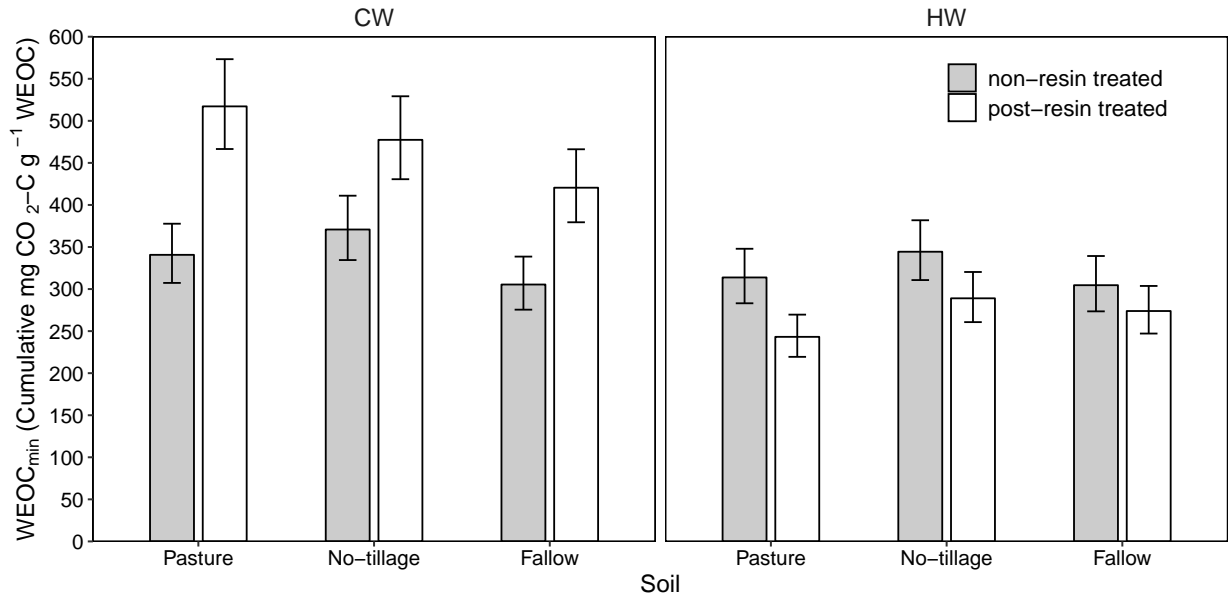
Land use did not affect WEOC<sub>min</sub> in either the CW or HW extracts ( $P = 0.129$ , Fig. 4.1). In the original (non-resin treated) extracts there was no difference ( $P > 0.05$ ) in WEOC<sub>min</sub> between CW and HW extracts but, after resin treatment, WEOC<sub>min</sub> was greater in CW (mean 47%) compared to HW extracts (mean 27%,  $P = 0.003$ ). There was a tendency for WEOC<sub>min</sub> to be higher in the CW post-resin treated sub-fractions of pasture and no-tillage soils than in the fallow soil. These changes in WEOC<sub>min</sub> were not observed in the HW post-resin treated sub-fractions.

The WEOC utilization followed a pattern consistent with WEOC<sub>min</sub> ( $r = 0.88$ ,  $P < 0.001$ ), with the highest WEOC utilization ( $\sim 70$ –84% for fallow and pasture, respectively) measured in the CW post-resin treated sub-fraction. In addition, WEON utilization was greater but proportional to the WEOC utilization ( $r = 0.85$  and 0.41 for CW and HW, respectively;  $P < 0.001$ ), mainly in the CW post-resin treated sub-fraction (see Fig. B.6 in Appendix B). However, there was no evidence of a significant land use effect for any WEOM fraction and/or sub-fraction ( $P = 0.239$ ).

Despite differences in WEOC<sub>min</sub> between the CW and HW post-resin treated sub-fractions, the proportion of CO<sub>2</sub>-C produced per unit of WEOC utilized averaged 58 to 62%, regardless of land use, extraction type or resin treatment.

### WEOM characteristics explaining variability in WEOC<sub>min</sub>

The results of the RF modeling showed that the combination of the nine quantitative and qualitative variables included in the model explained 75.8% of the variation in



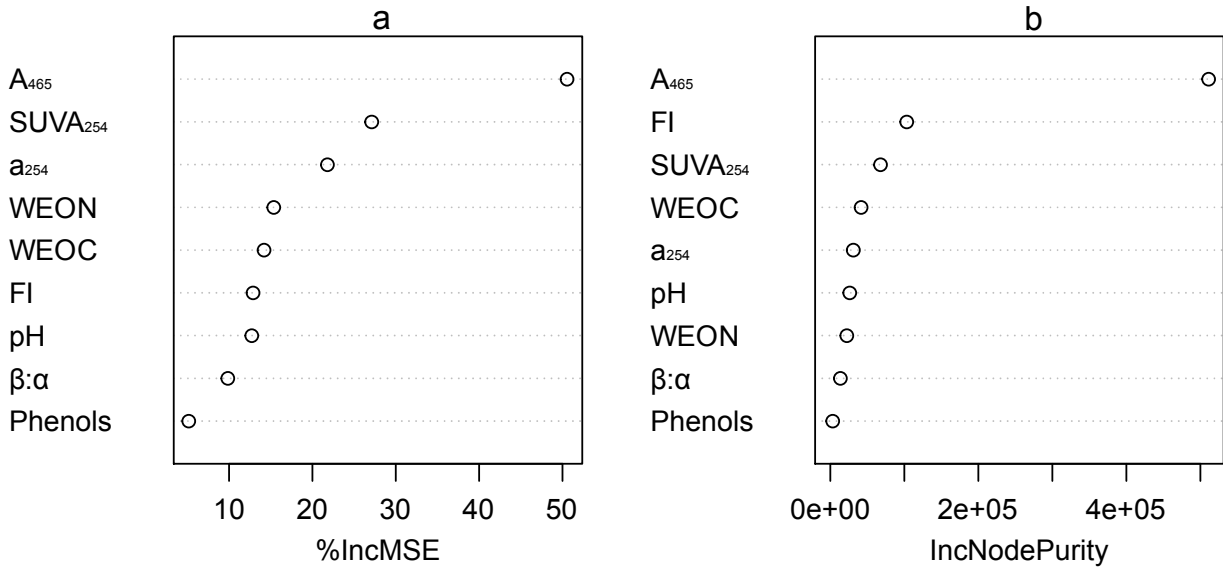
**Figure 4.1:** Cumulative WEOC mineralized ( $\text{WEOC}_{\min}$ ) over 14 days in non- and post-resin treated cold (CW) and hot water (HW) extracts of a soil under different land uses. Bars represent predicted means ( $n = 9$ ), error bars represent lower and upper confidence intervals.

the  $\text{WEOC}_{\min}$ . High values of IncMSE and IncNodePurity, corresponding to high variable importance (Kuhn et al., 2008), revealed that the  $A_{465}$  contributed most in explaining the changes observed in the  $\text{WEOC}_{\min}$  of the different pools and fractions (Fig. 4.2a and b). However, the regression between the  $A_{465}$  and  $\text{WEOC}_{\min}$  showed their poor relationship in the HW extracts (Fig. B.5 in Appendix B).

#### 4.1.4 Discussion

##### The land use effect on WEOM concentration

Soil surface general properties showed that the conversion of long-term pasture to cropping management reduced soil C and N contents (Table 4.1) and, consequently, the amount of WEOM per unit mass of soil. Previous studies have shown that the decrease in SOM content due to cultivation of former long-term pastures and native grasslands results from reduced belowground plant C inputs from annual arable crops rather than soil physical disturbance (Francis et al., 1999; Baker et al., 2007; Kätterer et al., 2011; Curtin et al., 2014). In addition, Sisti et al. (2004) and Raphael et al. (2016) stressed the importance of plant N inputs (biological fixation) on SOM(C) controls in rotational cropping systems under no-tillage management.



**Figure 4.2:** Outputs of the Random Forest algorithm describing the variable importance in explaining the WEOC mineralization ( $\text{WEOC}_{\min}$ ) response: a) Increased Mean square-error (%IncMSE) and b) Increased Node Impurity Index (IncNodePurity). The abbreviations, along with the corresponding units, of each descriptive variable are defined in Subsection 4.1.2 (Statistical analyses).

However, in the no-tillage soil, these inputs might have served as a nutrient source for N-demanding crops (wheat, barley), while the constant N inputs in the pasture soil would have maintained SOM levels. The results of the present study concur with these previous studies and align with the observed relative decrease in WEOM per unit of soil observed for both no-tillage and fallow soils with respect to the pasture soil, since none of the land use treatments underwent physical disturbance, but both cultivation and the complete absence of plants reduced soil WEOC and WEON. The larger decrease in WEOM than in total SOM ( $\sim 20$  and  $40\%$  for both TC and TN of no-tillage and fallow, respectively) also confirms the susceptibility of the soluble pool of SOM to changes in land use with respect to the bulk soil (Boyer and Groffman, 1996). In particular, the greater decrease in CWEOM with respect to HWEOM, for both no-tillage and fallow soils, indicates the larger turnover of the CW fraction with respect to the physically protected HW fraction, likely due to its active involvement in soil microbial processes and/or its dependency on fresh C inputs. Curtin et al. (2014) observed that the proportion of microbially mineralizable C in the pasture soil studied in the present work was considerably lower than the decrease in the C content following conversion to arable cropping, attributing their observations to an apparent underestimation of C mineralization by laboratory incubations. The results presented here do not disregard this hypothesis: however, they do suggest

that this observed decrease was subordinate to the contribution of plant inputs, and their ensuing role in the formation of the CW fraction (Uchida et al., 2012), not being replenished fast enough to compensate for losses due to mineralization of labile C forms from pasture soils.

Despite significant land use-induced changes in the absolute amount of WEOC, extractability of both CWEOC and HWEOC per unit of soil C was generally consistent with respect to different land uses (Table 4.3). A secondary effect of land use on WEOM formation was hypothesized, primarily regulated by SOM quality. For instance, the consistent release of WEOC from pasture and no-tillage soils suggests a relatively large presence of stable organic forms in the pasture soil, resulting in an equivalent contribution of less decomposed, plant-derived SOM, with respect to the no-tillage soil. Similarly, the tendency for the fallow soil to release less WEOC per unit soil C, although significantly less only in the CW extracts, suggests an enrichment in more complex and refractory SOM following progressive humification and depletion of labile components, and further indicates the crucial role of plant inputs on SOM quality and WEOC formation. These ideas are also supported by the more evident land use effect of the fallow soil on the WEON released per unit of soil N (reduction), since it is well established that plants and the larger microbial biomass associated with these, represent the major source of available N (Kögel-Knabner, 2002; Geisseler et al., 2010).

### Characteristics of WEOM fractions prior to resin treatment

The ratio between the CWEOC and HWEOC was consistent with previous studies that measured 2 to 7 times more HW than CW extractable C (Gregorich et al., 2003; Fernández-Romero et al., 2016), and confirms that the HW fraction represents the largest stock of soluble OC. The fact that this difference was larger in the no-tillage and pasture soils than in the fallow soil indicates the importance of plant inputs in the formation of the HWEOC fraction, but also suggests that HWEOC acts as a source of microbially available C in the absence of plants.

Across different land uses, the ~5 to 6-fold higher HWEON with respect to the CWEON indicated that the largest portion of available N is stored in the HW fraction (Rinot et al., 2018) and that this fraction is likely to be more N- than C-rich (3 to 4-fold WEOC increase). This assumption is somewhat confirmed by the lower WEOC:WEON ratios of the HW fraction (10–14) when compared to the CW fraction (14–18).

Variations in the fluorescence emission spectra at 465 nm excitation ( $A_{465}$ ) normalized for soil C, or determined on a standardized concentration, have previously been related to structural changes associated with the humification process (Milori et al., 2002, 2006). Since it has been well established that fluorescence intensities are a function of WEOC concentration (Mobed et al., 1996; Rosa et al., 2005), this



study investigated the integrated areas under such emission spectra normalized for the WEOC concentration of the sample.

The higher  $A_{465}$  in the CW extracts compared to the HW extracts indicate that more decomposed and humified material was released with the CW extraction than with the HW extraction. This is in accordance with previous studies relating the source of the mineral-associated WEOM (Kleber et al., 2007) to organic material that has undergone preferential microbial fractionation, structural rearrangement and/or *de novo* synthesis, termed “soil humus” (Gregorich et al., 2000).

According to Fuentes et al. (2006), the  $A_{465}$  effectively distinguishes the humic features from other WEOM components. The region of excitation at 415–470 nm has been associated with molecular entities with extended  $\pi$ -systems, such as those resulting from enhanced structural condensation, conjugation and heteroatom (e.g. -O and -N) substitution of quinone-like structures (Ghosh and Schnitzer, 1980) and semiquinone free radicals (Senesi, 1990; Rosa et al., 2005) originating from lignin decomposition (Senesi, 1990; Cory and McKnight, 2005) and, specifically, phenol oxidation (Zech et al., 1997). These were shown to exhibit the characteristic fluorescence shift towards longer emission wavelengths (Fellman et al., 2010), hence contributing to the fluorescence at 465 nm excitation, they are likely to be related to the increase in the degree of humification (Milorí et al., 2002) of CW extracts. The HW fraction had, in fact, a greater content of phenolic C, but lower  $A_{465}$ .

Earlier work on soil humic substances demonstrated a strong relationship between structural moieties fluorescing at long wavelengths (red shift), with low intensity, and linearly condensed, substituted aromatic rings and other highly unsaturated, conjugated systems (Peuravuori et al., 2002), the latter correlated with measures of  $SUVA_{254}$  (Traina et al., 1990; Novak et al., 1992; Weishaar et al., 2003). For the non-resin treated extracts, values of  $SUVA_{254}$  and  $A_{465}$  indicated that the bulk of both CWEOM and HWEOM was made of aromatic moieties characterized by a relatively early stage of humification. The  $SUVA_{254}$  values of CW and HW extracts are, in fact, comparable with those reported for fulvic acids (Rodríguez and Núñez, 2011), which generally display lower  $SUVA_{254}$  than humic acids due to their lower unsaturation degree and larger content of carboxylic groups, conferring them greater solubility in water. Of note is the increase in  $SUVA_{254}$  in the CW with respect to the HW extracts, which suggests that the formation of the readily soluble OM fraction resulted from the progressive oxidation, condensation and polymerization of structurally less complex WEOM units, the latter being more abundant in the physically protected HW fraction.

The FI and  $\beta:\alpha$  indicated that the largest proportion of both CWEOM and HWEOM derived from SOM and plant decomposition, and confirmed that WEOM with a larger humification degree and aromatic C content is released from highly decomposed SOM (McKnight et al., 2001; Cory and McKnight, 2005; Kim et al.,

2006; Rodríguez et al., 2014; Fernández-Romero et al., 2016).

According to McKnight et al. (2001), values for FI reported in this study would correspond to fulvic acids, since humic acids generally exhibit FI values  $<1.2$  (Chen et al., 2003; Rodríguez et al., 2014). Differences in  $FI \geq 0.1$  units have been related to significant differences in aromaticity (Kim et al., 2006; Rodríguez et al., 2014). Thus, the FI increase  $>0.1$  in the HW fraction with respect to the CW fraction is indicative of less condensed aromatic structures or conjugated aliphatic polymers (Senesi et al., 1991), confirmed by the corresponding lower  $SUVA_{254}$  values.

Previous studies reported a strong correlation of HWEOM with microbially derived polysaccharides, concluding that this fraction was of microbial origin (Haynes and Francis, 1993; Ghani et al., 2003). Polysaccharides were not specifically assessed in the current study; however, while polysaccharides of microbial origin might constitute a portion of the HWEOM fraction, these may have represented only a limited proportion of the HWEOM in this study. It is likely, however, that the HW fraction reflects an early stage of decomposition and humification, given the values of  $A_{465}$ ,  $SUVA_{254}$  and FI previously discussed.

In contrast, the increase in aromaticity with increases in HWEOM reported by Fernández-Romero et al. (2016) does not align with the  $SUVA_{254}$  values reported in the current study: however, their results may be explained by the larger total phenolic concentration observed in the HW extracts of the current study. In the HW extracts, the greater proportion of phenolic C and the lower  $SUVA_{254}$  with respect to the CW extracts suggest that the aromaticity is not solely derived from phenolic compounds. However, it is likely that the phenolic C in the HW extracts derives from lignin decomposition (Senesi et al., 1991), as confirmed by the low values for FI and  $\beta:\alpha$ . Advanced lignin decomposition, involving the progressive oxidation of aliphatic chains of phenolic moieties to carboxyl groups, leads to decreased aromaticity and increased solubility.

#### Land use effect on the characteristics of WEOM prior to resin treatment

For both CW and HW extracts, patterns for  $A_{465}$ , progressively increasing from no-tillage to fallow soils, indicated a larger accumulation of humic material in absence of fresh plant inputs in the soil (Barré et al., 2010; Shiau et al., 2017), but also an increased turnover of humified material (priming effect) resulting from the larger removal of plants in a cropping system or the enrichment in less humified SOM forms from residue decomposition (Bayer et al., 2000; Mazzilli et al., 2014). The latter would also explain the tendency for  $SUVA_{254}$  values to decrease in the no-tillage soil, indicating an increase of the aliphatic character of WEOM. Despite the decrease in WEOC concentration and absorbance, the consistent values for  $SUVA_{254}$  in the CW extracts of the fallow soil with respect to the pasture soil revealed that the relative proportion of aromatics remained constant, possibly due to the depletion of

both aliphatic and aromatic/phenolic C after long-term fallow. Values of  $SUVA_{254}$  in the HW extracts paralleled the decrease of  $a_{254}$  and the phenolic content from pasture to fallow, suggesting that, in the HW extracts,  $SUVA_{254}$  is affected by the progressive depletion/humification of phenolic moieties in absence of plants, as well as by the quality of the plant inputs (i.e. an enrichment of non-absorbing, aliphatic compounds).

#### Resin treatment effects on WEOM uptake and characteristics

The strong linear relationship between the WEOC retained by the resin and the chloride released indicates that anion exchange was the predominant mechanism of WEOM removal. However, the fact that, for both CW and HW extracts, the slopes of this relationship were  $<1$  indicates that other retention mechanisms, such as surface adsorption (Davies and Thomas, 1951; Fu and Symons, 1990), might have occurred concomitantly (Croué et al., 1999), particularly in the HW extracts. Competition from other anions with the WEOC for the resin binding sites was unlikely to be significant (Appendix B, Table B.1a). Hence, considering that the HW extracts contained significantly more phenolic C than the CW extracts and that, for both extractions, phenols were not detected in the post-resin treated sub-fraction, it implies that phenolic moieties, non-ionized at the pH measured for the CW and HW fractions (4.5–6), interacted with the inner surface of the resin (Fu and Symons, 1990). This assumption is corroborated by earlier studies, which demonstrated that phenol removal by macroporous anion exchange (Carmona et al., 2006; Caetano et al., 2009) occurs mainly by hydrophobic interaction via van der Waals adsorption (Tamamushi and Tamaki, 1959; Kunin and Suffet, 1980; Aceto et al., 1995; Ioannidis and Anderko, 2001). Furthermore, it is unlikely that phenols would have reacted with the resin chloride, since such a reaction would have caused a greater release of  $H^+$ , and the decrease in pH observed in the HW extracts was smaller than in that in the CW extracts. On the other hand, the stoichiometry of CWEOC exchange for chloride ions and the higher charge density of the CWEOC with respect to the HWEOC also suggest that the CWEOM had a greater hydrophilic character and content of oxygenated, mainly carboxylic functional groups (Croué et al., 1999), effectively displacing chloride from the resin surface. Nevertheless, the WEOC uptake efficiency was  $\sim 20\%$  greater for the HW than for the CW extracts, despite the relatively higher concentration of HWEOC with respect to the CWEOC, indicating a greater adsorption capacity of HWEOC per unit of resin and ruling out any resin saturation effect after equilibration of the CWEOC (Appendix B, Fig. B.3). Thus, the greater removal of HWEOC with respect to the CWEOC appears to be the result of co-existing mechanisms of ion exchange and hydrophobic interaction, withdrawing phenolic moieties and allowing for a clear-cut separation of an uncharged/protonated, phenolic-free sub-fraction.

Further evidence for both charge and quality (functional group) differences between CWEOM and HWEOM is provided by the greater decrease in the pH of the CW extracts compared to that observed for the HW extracts. Since a relatively large proportion of the carboxylic groups is most likely dissociated at the pH of the CW extracts (5.5–6), their uptake by the resin and the concurrent release of chloride ions will lead to an increase of the  $H^+$  concentration, thus significantly decreasing the pH.

For both CW and HW extracts, the general decrease in values of  $SUVA_{254}$ ,  $a_{254}$ ,  $A_{465}$ , FI and  $\beta:\alpha$  following the resin treatment indicates that the anion exchange resin consistently isolated an aliphatic, non-humified WEOM sub-fraction, likely derived from more recent, microbial outputs. Despite this general conclusion about the characteristics of only a small proportion of the bulk CWEOM and HWEOM, identified with the post-resin treated sub-fraction, the patterns revealed by FI and  $a_{254}$  suggest an influence of plant and SOM inputs on microbial outputs, such that their turnover is lower in the HW sub-fraction, and in particular in the pasture soil, whilst the microbial signature is greater in the CW post-resin treated sub-fraction.

### **Land use effect on the characteristics of WEOM following resin treatment**

In contrast to the non-resin treated CW and HW fractions, the post-resin treated sub-fraction reflected differences in extractability of CWEOC and HWEOC, as revealed by the poorer correlation between CWEOC and HWEOC in this sub-fraction. This was due to the larger spatial variability of CWEOC observed for the different land uses (mainly fallow and pasture soils) with respect to the HWEOC (data not shown, included in the REML LMM), indicating a higher sensitivity of the CWEOC, and in particular of its post-resin treated sub-fraction, to micro-scale changes in land use. However, the consistent WEOC and WEON extractability of the post-resin treated CW and HW sub-fractions suggest the coupling of WEOC and WEON in the organic moieties remaining following the resin treatment.

### **The importance of WEOM properties in explaining WEOC biodegradability patterns**

Potentially mineralizable C bioassays are commonly performed to estimate the labile OM fraction (Haynes, 2005; Hopkins, 2008). Here this fraction was assessed by determining the amount of  $CO_2$ -C released per unit of WEOC in CW and HW extracts with and without anion exchange resin treatment. The short-term incubation assay results presented in this study were consistent with previous laboratory studies reporting that 10–44% of the water-soluble organic C decomposes in a relatively short time frame (Kalbitz et al., 2000; Gregorich et al., 2003; Kalbitz et al., 2003; Fellman et al., 2008).

Hot water WEOC has been associated with the labile and readily mineralizable OM fraction (Ghani et al., 2003; Gregorich et al., 2003) due to its high correlation with microbial activity, mineralizable N and total carbohydrate content (Leinweber et al., 1995). As such, it has been regarded as being susceptible to land use changes. However, contrarily to expectations,  $WEOC_{min}$  did not differ with either CW and HW extracts or land use. This raised the question as to whether there is a sub-fraction of both the CWEOM and HWEOM that determines the variation in biodegradation and  $WEOC_{min}$  and its responsiveness to the land use effect.

The higher  $WEOC_{min}$  in the CW post-resin treated sub-fraction with respect to the corresponding non-resin treated fraction shows that the resin removed the less biodegradable fraction of WEOC. In contrast, these changes were not observed between the non-resin treated and post-resin treated sub-fractions of HW extracts.

Therefore, the importance of the measured WEOM properties in explaining and predicting  $WEOC_{min}$  was further investigated. It must be stated that the study was not designed for a full RF implementation, with training and validation, since the variables and  $WEOC_{min}$  were results of the experimental factors. However, the algorithm showed strong evidence of a relationship between the measured variables and  $WEOC_{min}$ .

The RF model revealed the  $A_{465}$  (Milori et al., 2002) to be an important variable in explaining the  $WEOC_{min}$  response (Fig. 4.2). Its strong correlation with  $SUVA_{254}$  suggests, furthermore, that changes in  $WEOC_{min}$  are strongly related to changes in the content of aromatic structures. Previous studies have shown that less humified material has shorter turnover times with respect to more humified material with a larger degree of aromaticity (Qualls and Haines, 1992; Almendros and Dorado, 1999; Kalbitz et al., 2003; Fellman et al., 2008). Thus, it may be concluded that in presence of more complex, humified material, microorganisms preferentially utilize the aliphatic WEOM fraction; this would suggest that  $WEOC_{min}$  in the CW extracts is largely driven by the  $WEOC_{min}$  of the post-resin treated sub-fraction. However, while differences in the humification degree might explain the variation in the  $WEOC_{min}$  response between the non- and post-resin treated CW extracts, they do not explain the  $WEOC_{min}$  pattern observed in the HW extracts, where large changes in  $A_{465}$  were associated with comparable  $WEOC_{min}$  (Appendix B, Fig. B.5). In addition, since the HW fraction was characterized by lower  $A_{465}$  and  $SUVA_{254}$  with respect to the CW fraction, it would have been expected to observe larger values for  $WEOC_{min}$ , but this was not the case.

The RF algorithm indicated that the concentrations of WEOC and WEON are a weak predictor of  $WEOC_{min}$  (Lundquist et al., 1999), regardless of extracted fraction, land use or resin treatment (Fig. 4.2a). In fact, large differences in the size of the CW and HW fractions corresponded to a comparable  $WEOC_{min}$  (Fig. 4.1). On the contrary, small differences in WEOC and WEON concentrations of

the post-resin treated sub-fraction resulted in a significantly different  $WEOC_{min}$  (Fig. 4.1). This is in contrast with the idea that the quantity of available substrate determines WEOM decomposition rates (Uchida et al., 2012). Furthermore, the lower  $WEOC_{min}$  of the non-resin treated CW fraction with respect to the CW post-resin treated sub-fraction, containing only a small proportion of the total C $WEOC$  and C $WEON$ , suggests that the quantity of WEOM is not a limiting factor for  $WEOC$  biodegradability (see also Chapter 3). This hypothesis is further supported by the lack of any significant relationship between the concentrations of H $WEOC$  and H $WEON$  and their biodegradability.

Although the other variables were given a lower importance rank by the RF model, they still contributed in explaining the variation in  $WEOC_{min}$ . This suggests that a combination of variables, that relate to the composition and source of the H $WEOM$ , may clarify the  $WEOC_{min}$  response in HW extracts.

A large proportion of the H $WEOC$  in this study comprised phenols, which are not a preferentially utilizable C source for microorganisms relative to carbohydrates and proteins (Rosa et al., 2005), although decomposition rates for phenolic compounds have been shown to differ (Qualls and Haines, 1992). In contrast, it is likely that amino acids and (plant-derived) short-chain oligopeptides also largely contributed to the H $WEOM$ , given the low C:N ratios, the large proportion of  $WEON$  comprising the TDN ( $\sim 86\%$ ) and the polar nature of  $WEON$ , inferred by its large decrease after resin treatment (83–87%). Previous studies have shown that such compounds, and in particular peptides (Miltner et al., 2009), can be taken up intact (Higgins and Payne, 1980) and readily utilized by microbes (Farrell et al., 2013). Farrell et al. (2014) also recently demonstrated that the soil microbial uptake of peptides is driven by C rather than N excess: it becomes, therefore, intuitive to conclude that, in the current study, microorganisms preferentially utilized the H $WEON$  pool, rather than phenols, as a C source to satisfy their metabolic needs. As a result, the excess N was released as  $NH_4^+$  (Gregorich et al., 2003; Farrell et al., 2014; Mooshammer et al., 2014) (Appendix B, Fig. B.4). However, it is also possible that some  $WEOC_{min}$  derived from phenols and other WEOM sources (carbohydrates, amino-sugars, fulvic acids), including the constituents of the post-resin treated sub-fraction. We also cannot exclude, albeit minor, an inhibitory effect of some phenolic compounds, such as tannins and flavonoids, on H $WEOM$  decomposition. These findings may partly contribute to explain the comparable (and lower than expected)  $WEOC_{min}$  of the HW extracts with respect to the CW extracts. However, they contradict the idea that H $WEOM$  represents the most labile and bioavailable OM source.

Reasons for the large differences in  $WEOC_{min}$  between CW and HW post-resin treated sub-fractions are not intuitive, since the resin treatment consistently isolated an uncharged/protonated, aliphatic, non-humified and likely microbially derived

WEOM sub-fraction. The fact that  $\text{WEOC}_{\min}$  of this functionally and biochemically different sub-fraction appeared to be affected by land use (pasture vs. fallow) in the CW but not in the HW extracts suggests that these microbial outputs and their mineralization potential are largely affected by the quality of plant inputs.

The large proportion of WEON comprising the TDN of the post-resin treated CW sub-fraction and the relatively proportional amount of WEOC and WEON utilized in this fraction corroborate the coupling of WEOC and WEON in the post-resin treated sub-fraction previously suggested. In the post-resin treated HW sub-fraction, the lower proportion of WEON in TDN, shared with  $\text{NH}_4^+$ , which remains unaltered after the incubation, and the relatively greater utilization of WEON compared to WEOC suggest that the available C source is not easily degradable and indicates the utilization of the N-containing organic C sources, in this case, likely for both its C and N (Mooshammer et al., 2014), ruling out any influence of  $\text{NH}_4^+$  on either C or N utilization (Gregorich et al., 2003). Further work should therefore (i) examine the effect of complex substrate decomposition (e.g. plant litter) on microbial synthesis of low molecular weight organic neutrals, such as uncharged (amino) sugars, amino acids and oligopeptides, and their role in  $\text{WEOC}_{\min}$  of the CW fraction and (ii) investigate the role of (N-containing) amphiphilic biomarkers (e.g. lipids) in the  $\text{WEOC}_{\min}$  of the HW fraction.

Interestingly, regardless of differences in  $\text{WEOC}_{\min}$  revealed by the post-resin treated sub-fraction, the proportion of  $\text{CO}_2$  produced per unit of WEOC utilized (referred to as metabolic efficiency) was the same for all fractions and sub-fractions. Such an observation indicates that the WEOC mineralization is a function of a metabolic threshold, likely to be governed by substrate composition (Sugai and Schimel, 1993) and, thus, nutrient limitation.

#### 4.1.5 Conclusions

The anion exchange resin treatment separated two functionally and biochemically different fractions of WEOM. While, in the CW extracts, the post-resin treated sub-fraction exhibited greater biodegradability compared to the bulk extract, the mineralization response of this sub-fraction in the HW extracts was comparable with the bulk extract. The information revealed by the fluorescence indices was in agreement with that of the UV spectroscopy revealing that anion exchange consistently isolated an aliphatic, non-humified WEOM sub-fraction, likely derived from more recent, microbial outputs, as opposed to the aromatic nature of the negatively charged and/or hydrophobic fraction of the WEOM retained by the resin, characterized by a relatively early stage of humification of SOM and plant components. Further evaluation is, however, needed to substantiate the role of hydrophobic interaction in WEOM adsorption by anion exchange resins in order to further develop chemical and biochemical characterization of WEOM fractions.

The RF model elucidated the complex relationship between the quantitative and qualitative properties of WEOM and its mineralization potential. The model identified the  $A_{465}$  humification index (Milorí et al., 2002) normalized per unit of WEOC as an important variable explaining the biodegradability patterns of WEOM extracted from different SOM fractions and isolated with anion exchange resin treatment. The utilization and further validation of the  $A_{465}$  as a simple but robust predictor of WEOM mineralization in modeling studies is, therefore, proposed. However, the results indicated that in presence of more complex, humified material, the mineralization of the aliphatic WEOM fraction makes a significant contribution to the overall mineralization response of the CWEOC, but not the HWEOC. Thus, additional factors, such as the relative proportion of phenolic compounds and labile, SOM or microbially derived N-containing compounds, can also control this response, as evidenced by the increase in  $\text{NH}_4^+$  following mineralization of C in the HW extracts. This reveals the importance of plant inputs, their decomposition, and the allocation of the ensuing compounds on microbial metabolism and the formation of WEOM sources and sinks.

## Supplementary material

Supplementary information to this Chapter, including the field trial design (Fig. B.8), the determination of the resin treatment equilibration times (Section B.1) and the calculation and evaluation of the  $A_{465}$  humification index (Section B.2), is included in Appendix B.

## 4.2 The effect of pH on the retention efficacy of WEOM by anion exchange resin

### 4.2.1 Rationale

We previously observed that the pH of DOM and WEOM varies in soils under different land use, during leaching, or in cold (CW) or hot water (HW) fractions of soil extracts (Chapter 3 and Section 4.1). Since DOM/WEOM represents a heterogeneous assembly of a wide spectrum of organic molecules acting like weak-acid polyelectrolytes (Stevenson, 1994), these changes in pH might alter the degree of dissociation of ionizable functional groups ( $-\text{COOH}$ ,  $-\text{OH}$ ,  $=\text{NH}$ ) and/or affect keto-enol tautomerism, thus impacting the resin retention efficacy by anion exchange from soil solutions.

Earlier studies examining the uptake of solution poly-electrolytes (Stevenson, 1994; Ríos and Urzúa, 2001; Li and SenGupta, 2004) by means of anion exchange showed that the uptake mechanism (ion exchange, hydrophobic interaction or their



combination) and the stoichiometry of exchange highly depend on the solution pH (Croué et al., 1999; Ioannidis and Anderko, 2001; Kanazawa et al., 2004; Ku et al., 2004; Caetano et al., 2009). It becomes, therefore, important to determine the effect of the soil solution pH on the retention of DOM/WEOM on anion exchange resins, in order to maximize resin separation efficacy and introduce a standardized fractionation method for the study of WEOM biodegradability.

The objectives of this work were to: i) assess the effect of different pH conditions (commonly found in agricultural soils) on WEOM retention by anion exchange resin (by altering the pH of soil extracts); and ii) determine the optimal resin:extract ratio for the maximal WEOM retention/fractionation efficiency. These objectives were, thus, pursued before the experiment presented in the previous Section.

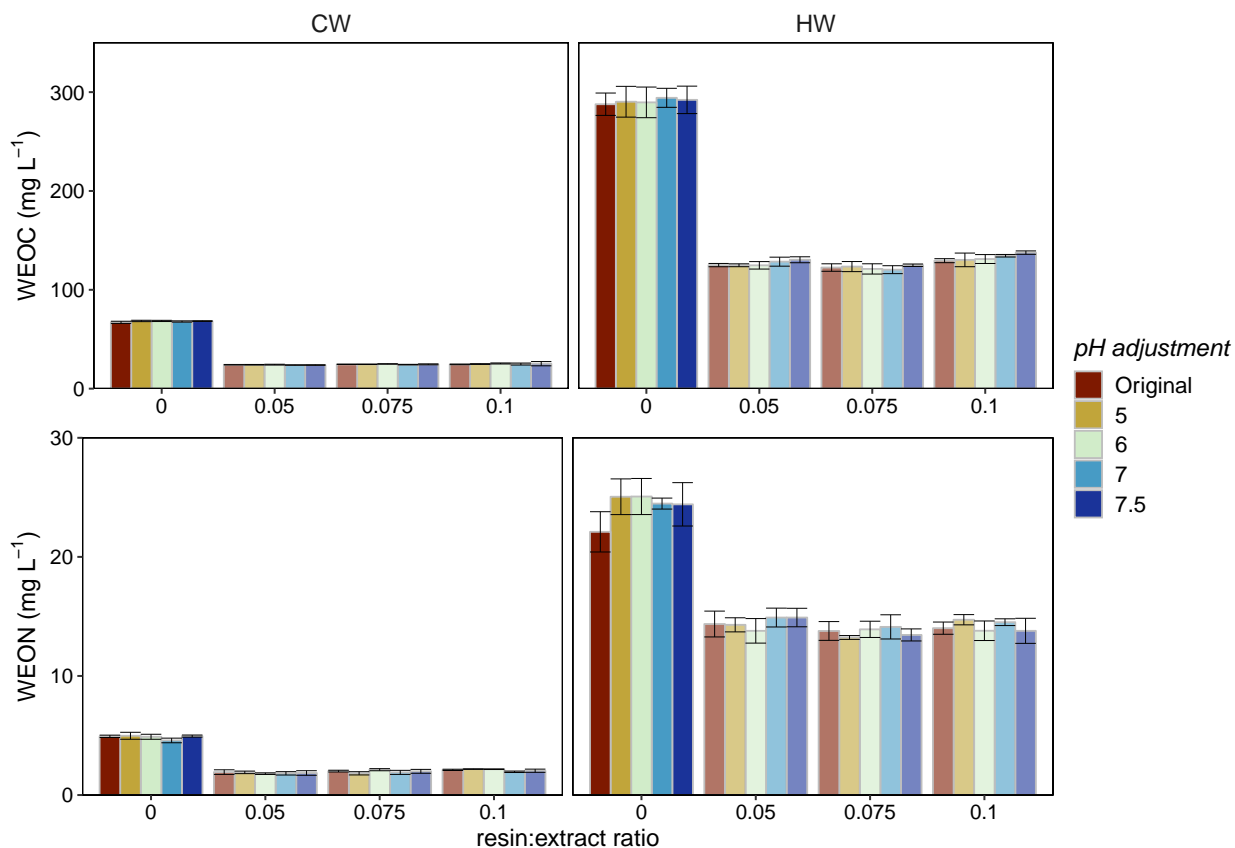
## 4.2.2 Experimental

The soil selected for this experiment (40.9% sand, 50.2% silt, 8.9% clay), classified as allophanic in the New Zealand soil classification (Hewitt, 2010), was collected from a paddock under dairy farming in the Waikato region of the North Island, New Zealand<sup>1</sup> (38° S and 175° E; annual rainfall between 1200–1400 mm; mean soil temperature at 1 m depth of 13–15°C). The main selection criteria were a high SOM content (TC 8.8%, TN 0.82%) and a low pH (5.52). Soil properties were determined following the procedures and standard protocols described in Subsections 3.2.1 and 4.1.2.

Air-dried (9.2%  $\theta_g$ ) and sieved (<4 mm) soil subsamples were extracted in triplicate, as described in Subsection 4.1.2. Briefly, soils were extracted with deionized water for 1 h at 20°C (CW extraction) and subsequently at 80°C (HW extraction) for 16 h at a soil:water ratio of 1:10. The extracts were centrifuged at 2831 *g* and then at 17 696 *g* for 5 min, respectively, and supernatants were filtered through 0.45- $\mu$ m PVDF filter membranes. Each CW and HW extract was divided into five subsamples, either kept at its original pH value (5.7 and 5.6 for CW and HW, respectively) or adjusted to pH 5.0, 6.0, 7.0 and 7.5 using 0.05 M HCl/0.05 M NaOH for the CW extracts and 0.1 M HCl/0.1 M NaOH for the HW extracts. An aliquot of 100 mL of each extract at the designated pH level was added with a macroporous, crosslinked polyacrylamide, strong base (Type I) chloride form anion exchange resin (Lanxess Ionac MacroT)<sup>2</sup> at a ratio of 1, 1.5 and 2 g resin/20 mL solution and shaken for 2 h. Samples were then filtered through GF-A syringe filters to separate the extracts from the resin. Resin-treated and original samples were immediately assessed for pH, NPOC, TDN and mineral N concentrations, and UV

<sup>1</sup>This soil belongs to the Plant and Food Research ‘N Min’ archive.

<sup>2</sup>The pH values of the solutions were within the operating pH range of the anion exchange resin (0–14).



**Figure 4.3:** Concentrations of WEOC and WEON in cold (CW) and hot water (HW) extracts at their original or adjusted value of pH before (resin:extract ratio = 0) and after the resin treatment (resin:extract ratio = 0.05 (1:20), 0.075 (1.5:20) and 0.1 (2:20)). Bars are means ( $n = 3$ )  $\pm$  s.d. WEOC l.s.d. ( $\alpha = 0.05$ ) = 2.1 and 25.5 mg L<sup>-1</sup> and WEON l.s.d. ( $\alpha = 0.05$ ) = 0.47 and 3.13 mg L<sup>-1</sup> for CW and HW, respectively.

absorbance at 254 nm (details in Subsection 4.1.2). All analyses were done within 24 h.

Data were analyzed using analysis of variance (ANOVA), treatment effects were compared by means of least significant differences (l.s.d.) at the 5% level of significance and multiple comparisons were expressed by adjusted  $P$ -values using the less conservative, family-wise Holm correction (Holm, 1979), valid under arbitrary assumptions. Statistical analyses and graphs were done in R version 3.5.1 (R Core Team, 2018).

### 4.2.3 Results and discussion

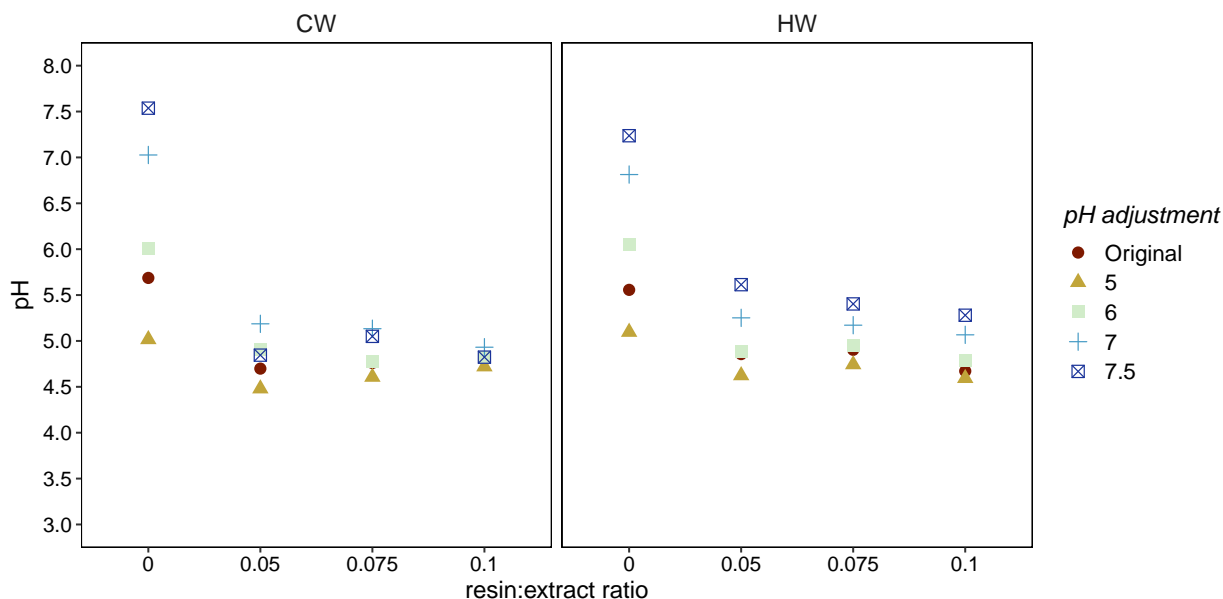
#### Changes in WEOC and WEON concentration

The retention of WEOC by anion exchange resin was not affected by pH differences within the range 5.0–7.5 ( $P = 0.28$  and  $0.96$  for CW and HW, respectively). The resin treatment removed 63–65% of the initial WEOC in the CW extracts (CWEOC) and 54–58% of WEOC in the HW extracts (HWEOC), regardless of pH and resin:extract ratio (Fig. 4.3). Differences in pH did not alter the WEOC concentration of either non- or post-resin treated, CW and HW extracts, at any resin:extract ratio ( $P = 0.68$  and  $0.1$  for CW and HW, respectively). This would allude to a consistent degree of WEOC dissociation at pH 5.0 and above, and disregard resin saturation effects for lower resin:extract ratios.

The fact that there remained a significant proportion of WEOM after the resin treatment of both CW and HW extracts at all resin:extract ratios suggests an incomplete removal due to chemical properties of WEOM. Similar observations have been reported by Boyer et al. (2008), who attributed least adsorption for microbial humic substances. It is noteworthy that the retention of CWEOC was consistent with that observed for the soil type and the different land uses discussed in Section 4.1, suggesting a relatively constant proportion of aromatic and/or conjugated, weakly acidic CWEOM compounds, regardless of higher concentrations. Certainly, the reduced adsorption of HWEOC indicated longer equilibration times for the HWEOC fraction with respect to CWEOC, as shown in Fig. B.9 and, subsequently, in the results presented in Section 4.1. These further method adjustments excluded the likelihood of resin saturation after equilibration of the CWEOC.

Similarly, there was no evidence to suggest a significant effect of pH changes on the resin retention of CWEON ( $P = 0.16$ ), albeit there was some evidence of a pH effect on the uptake of HWEON ( $P = 0.03$ ). However, this effect was mainly caused by an initially lower, yet not significant, HWEON concentration (Fig. 4.3), leading to a lower amount of retained WEON moieties. It is assumed that this lower HWEON concentration of the original HW samples is an artefact caused by instrumental drift or larger variation due to sequential extraction, not changes in pH (the concentrations at pH 5.0 and 6.0 were consistent). In fact, as for WEOC, altering the pH did not cause significant changes in WEON concentration of either non- or post-resin treated extracts at any resin:extract ratio ( $P = 0.65$  and  $0.73$  for CW and HW, respectively). This suggested a consistent WEON retention performed by the resin.

Regardless of pH and resin:extract ratio, the resin treatment removed a consistently larger proportion of CWEON (57–62%) with respect to the HWEON (41–44%). This is likely due to the longer equilibration times needed for the HWEOC fraction previously discussed, but cannot exclude substantial qualitative

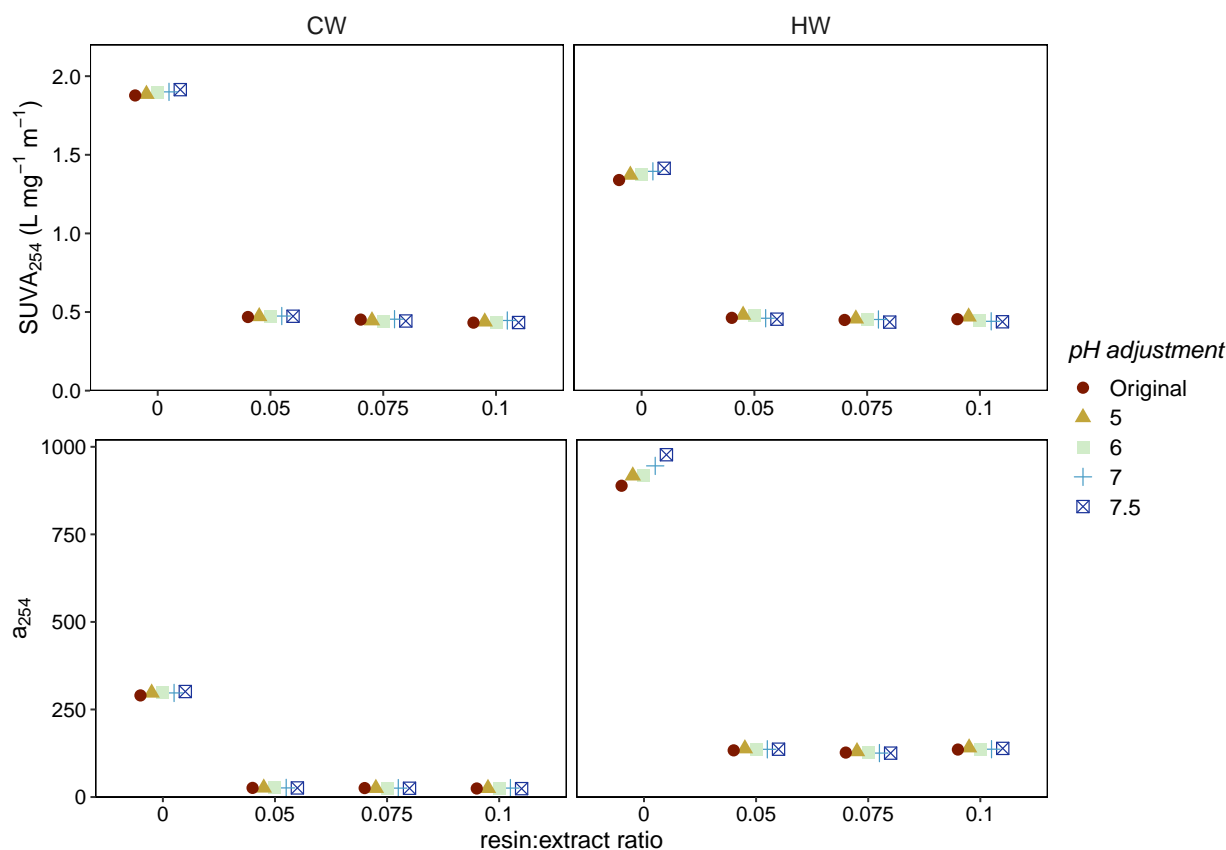


**Figure 4.4:** Changes in pH values of cold (CW) and hot water (HW) extracts following the resin treatment ( $n = 3$ ); resin:extract ratio = 0 (non-resin treated), 0.05 (1:20), 0.075 (1.5:20) and 0.1 (2:20). Error bars were omitted for clarity. l.s.d. ( $\alpha = 0.05$ ) = 0.54 and 0.34 for CW and HW, respectively.

differences between CW and HW extracts.

### Changes in pH

The resin treatment significantly reduced the pH of CW and HW extracts at any resin:extract ratio ( $P < 0.001$ ). The magnitude of this decrease was larger with increasing values of pH ( $P < 0.001$ ): the largest decrease occurred in the extracts with the highest pH values (Fig. 4.4), with an average of  $\sim 2.0$  to  $\sim 2.6$  and  $\sim 1.7$  to  $\sim 1.8$  pH units decrease for CW and HW extracts at pH 7.0 and 7.5, respectively. Such a decrease was not as pronounced for the extracts that were kept at their original pH values ( $\sim 0.7$ – $1$  pH unit decrease) or that were brought to pH 5.0 (up to 0.5 and 1 pH unit decrease for CW and HW, respectively). These results suggest a consistent degree of dissociation of the CWEOM in the range of pH values evaluated, such that the removal of a consistent number of negatively charged moieties (e.g. carboxylic groups) by the resin treatment, followed by the release of chloride ions, decreased the pH to the same level. Most (poly)carboxylic acids will be, in fact, dissociated in a pH range 4–6, common under forest and agricultural soils, in contrast to phenols, having a  $\text{pK}_a > 9$  (Carmona et al., 2006). The lower pH reduction in HW extracts was consistent with the lower proportion of HWEOM



**Figure 4.5:** Changes in SUVA<sub>254</sub> and a<sub>254</sub> values of cold (CW) and hot water (HW) extracts following the resin treatment ( $n = 3$ ); resin:extract ratio = 0 (non-resin treated), 0.05 (1:20), 0.075 (1.5:20) and 0.1 (2:20). Error bars were omitted for clarity. SUVA<sub>254</sub> l.s.d. ( $\alpha = 0.05$ ) = 0.05 and 0.1 L mg<sup>-1</sup> m<sup>-1</sup> and a<sub>254</sub> l.s.d. ( $\alpha = 0.05$ ) = 9.9 and 109 m<sup>-1</sup> for CW and HW, respectively.

taken up by the resin, suggesting that, at any given pH level within the examined range, there is a larger amount of WEOM capable of buffering the system, yet, a smaller amount of ionized moieties in HWEOM cannot be excluded.

For all ratios, the pH was reduced within a range of 4.5–5.2 and 4.6–5.6 for CW and HW extracts, respectively, indicating a greater variability in the pH values after the resin treatment of HW extracts (Fig. 4.4). There was tendency for pH to decrease with an increasing resin:extract ratio in the HW extracts, but these differences were not significant ( $P = 0.15$ ).

### Changes in aromaticity and/or conjugation

Values of  $SUVA_{254}$  and  $a_{254}$  differed between CW and HW non-treated extracts ( $P < 0.001$ , Fig. 4.5), but were not affected by changes in pH ( $P > 0.1$ ). This suggested that the optical properties of CW and HW extracts were not sensitive to alterations within the range of pH examined. While  $SUVA_{254}$  was lower in the HW extracts ( $\sim 1.4 \text{ L mg}^{-1} \text{ m}^{-1}$ ) compared to the CW extracts ( $\sim 1.9 \text{ L mg}^{-1} \text{ m}^{-1}$ ), an inverse pattern was observed for  $a_{254}$  ( $\sim 297 \text{ m}^{-1}$  and  $\sim 926 \text{ m}^{-1}$  for CW and HW, respectively), indicating that, despite the larger amount of absorbing moieties in the HW extracts,  $SUVA_{254}$  was largely affected by the relative concentration of non-absorbing moieties with respect to the CW extracts.

Both  $SUVA_{254}$  and  $a_{254}$  decreased significantly following the resin treatment of CW and HW extracts ( $P < 0.001$ ), and this decrease was not affected by differences in pH ( $P > 0.1$  for both  $SUVA_{254}$  and  $a_{254}$  in CW and HW extracts) or resin:extract ratio ( $P = 1$  for both  $SUVA_{254}$  and  $a_{254}$  and all extracts). Values of  $SUVA_{254}$  decreased below  $0.6 \text{ L mg}^{-1} \text{ m}^{-1}$ , regardless of extracted fraction, pH or resin:extract ratio (Fig. 4.5). Although  $a_{254}$  values after the resin treatment remained generally higher in the HW extracts than in the CW extracts ( $\sim 25 \text{ m}^{-1}$  and  $\sim 134 \text{ m}^{-1}$  for CW and HW, respectively), the relative decrease in  $a_{254}$  in the resin-treated HW extracts was lower than in the CW extract, resembling the patterns observed for the other variables. These results show that the bulk of CWEOM and HWEOM removed by the resin treatment consists of aromatic and/or conjugated compounds and is in line with the patterns observed from the soil under different land uses previously discussed (Section 4.1). Given the analogy, the chemical composition of the CW and HW fractions interacting with the resin might be comparable, but this can be confirmed only with further molecular and chemical characterization.

### 4.2.4 Conclusions

This work showed that changes in pH within the range 5.0–7.5 did not likely altering WEOM retention by anion exchange resin in both CW and HW extracts. It further allowed the suitable resin:extract ratio that enables the maximal WEOM retention without resin saturation to be determined. In addition, it showed that the capacity of the resin, at each resin:extract ratio examined, was not influenced by differences in concentration. The consistently lower WEOM retention in HW extracts suggested longer equilibration times should be considered. These have been, in fact, demonstrated and accounted for in the experiment described in the previous Section of this Chapter. The anion exchange resin allowed a clear separation of more aromatic and/or conjugated moieties from relatively more aliphatic compounds remaining after the resin treatment, which can be further investigated for their properties and biodegradability potential.

## Part II

# Molecular aspects of WEOM fractions in relation to their biodegradability

The previous Chapters emphasized the impact of WEOM composition, and in particular, qualifiers such as its degree of humification and aromaticity, on the ensuing biodegradability patterns. However, questions as to what causes inconsistencies among different WEOM fractions remain. The following Chapter aims to explain the observed results by expanding the understanding of the WEOM profile with powerful techniques such as fluorescence spectroscopy and high resolution liquid chromatography–mass spectrometry.



## Chapter 5

Searching for proxies of WEOM  
biodegradability: an integrated  
Fluorescence Data Mining and Liquid  
Chromatography–Mass Spectrometry  
approach

## Abstract

Understanding what chemical features determine the biodegradability of water extractable organic matter (WEOM) is important to optimize the use efficiency of its mineralization products and reduce losses through CO<sub>2</sub> emissions. The present study aimed to relate the molecular properties of WEOM fractions to their biodegradability response assessed by conventional laboratory bioassays. Cold and hot water soil extracts from different land uses were treated with anion exchange resin and separated according to their charge. Untreated and resin-treated samples were incubated for 14 days under controlled conditions to assess short-term WEOM biodegradation. Fluorescence excitation-emission matrices (EEMs) were unfolded by parallel factor analysis (PARAFAC) and functionally meaningful components capable of describing the variation in WEOM biodegradability and physicochemical characteristics were inferred using a data mining approach. The extracts were run through a hydrophilic interaction (HILIC) and a reversed-phase (C18) liquid chromatography and analyzed in a full scan mode on an Orbitrap mass spectrometer using heated electrospray ionization (HESI) in both positive and negative modes. Revealed compounds were classified according to a multidimensional stoichiometric constraints classification approach defined by Rivas-Ubach et al. (2018) or described by van Krevelen diagrams. Data mining of PARAFAC components identified six components capable of describing patterns of WEOM biodegradability and physicochemical properties. The van Krevelen diagrams showed an abundance of low molecular weight, N-containing compounds with variable unsaturation degree relative to whole extracts or resin-treated sub-fractions. These compounds were mainly identified as peptides or clustered over a region commonly defined as lignin-derived, which was only partly addressed by the classification applied. Non-targeted profiling revealed a significant contribution of quaternary ammonium compounds, which were attributed, along with small peptides, an important role in carbon mineralization. Principal component analysis showed structural differences among PARAFAC components, which were associated with the different WEOM fractions and sub-fractions and, thus, likely explain differences in WEOM mineralization.

**Keywords:** Biodegradation, Data mining, Excitation-emission matrices (EEMs), Nitrogen, LC-MS, PARAFAC, Water extractable organic matter.

## 5.1 Introduction

Small changes in the rate of soil organic matter (SOM) mineralization can potentially cause sizable changes in global carbon (C) dynamics. The latest estimates report CO<sub>2</sub> emissions from land use and management activities for the past decade

to equate  $1.5 \pm 0.7 \text{ GtC yr}^{-1}$ , an equivalent of  $\sim 13\%$  of total emissions comprising of both land use and fossil fuel emissions (Le Quéré et al., 2018). These contributed to an annual increase in the atmospheric  $\text{CO}_2$  concentration of  $4.7 \pm 0.02 \text{ GtC yr}^{-1}$ . But there remains considerable uncertainty ( $\sim 0.7 \text{ GtC yr}^{-1}$ ) due to limited understanding of the biogeochemical stoichiometries, analytical inconsistencies and model parametrization. Since the bulk of soil  $\text{CO}_2$  emissions comes from soil microbial activity, only a small portion of SOM is active at a time. This implies SOM is dissolved in order to be readily accessible and available for microbial utilization.

Dissolved organic matter (DOM) comprises  $<1\%$  of the SOM pool, but can be regarded as a “keyhole” for SOM mineralization processes. As a readily available C source for soil microorganisms, DOM can modulate soil microbial activity and community structure (De Nobili et al., 2001) and, thus, the stoichiometry between nutrient inputs and mineralization outputs. Plants and SOM are the major contributors to DOM (Gregorich et al., 2000; Kögel-Knabner, 2002; Malik and Gleixner, 2013), but biodegradation of DOM generates new DOM sources through heterotrophic production (Kallenbach et al., 2016). These cycles of DOM utilization and production depend on external conditions, such as changes in land use and vegetation cover, climate, soil physical properties and DOM’s initial composition (Chantigny, 2003).

The composition of DOM comprises a wide variety of low to high molecular weight compounds, with varying structure and complexity (Kalbitz et al., 2000), reflecting the quality of plant inputs and microbial processing (Guggenberger et al., 1994b; Malik and Gleixner, 2013). Common DOM constituents are carbohydrates, proteinaceous materials (proteins, peptides, amino acids) and other nitrogenous compounds (amides, amino sugars, nucleic acids), lipids, and a large component of aromatics and organic acids (Stevenson, 1994; Rumpel et al., 2004; Kögel-Knabner, 2006; Nannipieri and Eldor, 2009). In mineral soils, such heterogeneity is particularly accentuated in the A horizons, due to the proximity of plant litter and root exudates and fluctuations in abiotic controls (e.g. temperature and soil moisture/oxygen content). This leads to contrasting results and assumptions as to what chemical characteristics determine DOM biodegradability (Qualls, 2005).

The complexity and heterogeneity of DOM has led to a wide range of operationally defined fractionation approaches (Zsolnay, 2003). In this context, cold (C-) and hot (H-) water extractable organic matter (WEOM) fractions have been commonly studied as indicators of microbially available and biodegradable DOM (Marschner and Kalbitz, 2003; Haynes, 2005). While the CWEOM has been associated with humified SOM (Gregorich et al., 2000), particular focus has been given to the HWEOM fraction, due to its high correlation with the microbial biomass (Sparling et al., 1998; Ghani et al., 2003). Nevertheless, specific biomarkers (e.g. lignin-derivatives, polysaccharides) have been recorded to occur in both fractions

(e.g. Bu et al., 2010; Seifert et al., 2016), and their involvement in mineralization processes has been recently reevaluated (Kiem and Kögel-Knabner, 2003; Mueller et al., 2014). In this context, the study of the role of molecular classes and the dynamics of specific biomolecules during WEOM mineralization is fundamental for understanding and predicting the fate of soil C stocks and sources.

Even though DOM/WEOM biodegradability has been widely studied (Kalbitz et al., 2003), mainly in terms of cumulative mineralization fluxes, in-depth studies of its biodegradation and transformation at the molecular level are scarce, mainly due to its highly complex nature. Such information is, however, central to the understanding of DOM/WEOM formation, preservation and utilization.

Fluorescence and UV spectroscopic methods have been widely adopted in DOM studies (Senesi et al., 1991; Cory and McKnight, 2005; Embacher et al., 2008; Fernández-Romero et al., 2016), since optically active (chromophoric and fluorescing) moieties are characteristic features of DOM (Senesi et al., 1991; Leenheer and Croué, 2003). Their reactivity at specific wavelengths varies as a function of the molecular structure and complexity (Coble, 1996; Milori et al., 2002), thus enabling information about DOM/WEOM source and properties to be characterized. However, the need to link the large variety of chromophores (Korshin et al., 1997) and broadly defined fluorescence peaks from excitation-emission matrices (EEMs, Coble, 1996; McKnight et al., 2001) to more specific biochemical entities has paved the way for chemometric techniques, such as parallel factor analysis (PARAFAC), which is able to decompose the multiway EEM data array into independent components (Bro, 1997; Andersen and Bro, 2003).

Previous studies have proposed PARAFAC of EEM spectra for monitoring DOM/WEOM biodegradability (Fellman et al., 2008); however, contentions around the involvement of fluorescing components in DOM/WEOM biodegradation processes, as well as their relationship with chromophoric assemblies, including both aromatic and conjugated aliphatic structures, have not been resolved to date. The present work proposes a data mining approach to the study of biodegradability and optical properties of WEOM through PARAFAC components. The information acquired is, however, restricted to fluorescing moieties, and does not account for weakly or non-fluorescing material, such as cell lysates (Zsolnay et al., 1999), which may significantly impact the mineralization response. In addition, characteristic fluorescence components have been linked to molecular associations of different size, the latter being inversely proportional to their potential to be biodegraded. Thus, there is a need to complement information with other characterization techniques.

In recent years, the combination of chromatographic separation and high mass resolution and accuracy has been increasingly adopted to unravel the molecular characteristics of DOM (Sleighter and Hatcher, 2008; Cortés-Francisco and Caixach, 2013; Verkh et al., 2018). High-resolution liquid chromatography–mass spectrometry

try (LC–MS) is capable of resolving individual molecular identities out of complex mixtures, such as DOM/WEOM, and assigning them a molecular formula according to their mass-to-charge ( $m/z$ ) ratios. Yet, since DOM/WEOM is a heterogeneous association of organic moieties, more often than not, resolved species are not defined by a specific chemical identity. In spite of that, the calculation and the representation of O/C and H/C ratios (using van Krevelen diagrams) enables the biochemical classification of DOM/WEOM components (Kim et al., 2003; Roth et al., 2015; Mangal et al., 2016; Seifert et al., 2016; D’Andrilli et al., 2019) and informs us about the transformations following biodegradation (Bailey et al., 2017). However, in complex mixtures such as DOM/WEOM, different classes of compounds are often shown to overlap throughout such diagrams, hampering their correct interpretation and leading to inconsistent and inconclusive answers; thus, constraints of heteroatom to C molar ratios, atom counts and molecular weight have to be applied for a valid classification (Rivas-Ubach et al., 2018). Furthermore, such resolved formulae can result in a variety of different structural configurations, strongly defining the molecule’s functionality and, thus, its fate in biogeochemical cycles. In this context, (ring) double bond equivalents ((R)DBE) provide an estimate of the degree of unsaturation. Since double bonds with heteroatoms are included, normalized indicators, such as the ratio between DBE and C atoms (DBE/C), DBE minus oxygen atoms (DBE-O) and the aromaticity index (AI), have been additionally proposed (Koch et al., 2005; Hockaday et al., 2006; Koch and Dittmar, 2006; Cortés-Francisco and Caixach, 2013; Tfaily et al., 2013).

In the previous Chapter, CWEOM and HWEOM fractions were characterized and their sub-fractions were isolated following anion exchange treatment, in order to identify functionally and chemically different fractions of WEOM that could explain changes in biodegradability patterns. The extracts were obtained from a soil with contrasting management histories (long-term pasture to no-tillage cropping), including plots maintained without plant cover for 13 years. The degree of humification was inferred to be an important aspect affecting the mineralization of the CWEOM fraction, but not of the HWEOM fraction, conversely characterized by a large amount of phenols and organic N content. The results were consistent with previous studies hypothesizing a selective utilization of N-containing compounds in response to C limitation (McGill and Cole, 1981; Jones and Murphy, 2007; Farrell et al., 2014; Mooshammer et al., 2014). Furthermore, the data suggested a close coupling of C and N in WEOM moieties of the sub-fraction remaining following anion exchange treatment. It was hypothesized that they performed a regulatory role on the mineralization response as a function of secondary products from *de novo* synthesis or biodegradation processes in response to plant inputs. While the interdependence between C and N has been already recognized by earlier studies (Gärdenäs et al., 2011; Knicker, 2011), such a relationship and its bearing on the

overall mineralization response have not been investigated in WEOM sub-fractions isolated by anion exchange treatment. Such hypotheses, thus, require further insight in the biomolecular characteristics of these fractions and sub-fractions and their effect on microbial activity.

In the present work, an initial fingerprinting of WEOM by means of fluorescence EEMs was followed by PARAFAC data mining, for the selection of fluorescence components best describing biodegradability and selected biochemical properties, and WEOM profiling through a non-targeted LC–MS analysis (Halket et al., 2005). Ultimately, PARAFAC components were combined with LC–MS data of CWEOM and HWEOM fractions and anion exchange isolated sub-fractions (into a PCA) to relate the observed mineralization patterns with the optical and molecular properties of WEOM from a soil under different land uses.

## 5.2 Materials and methods

The treatments investigated in this study, including soils, extraction procedures, anion exchange fractionation details and related WEOC/WEON, optical and biodegradability measurements, were described in Section 4.1 of Chapter 4. They are repeated to some degree here for convenience.

### 5.2.1 Soils, extracts and WEOM fractionation

The soils investigated in this study were collected from a long-term field trial established in November 2000 near Lincoln, Canterbury, New Zealand (43° 40' S, 172° 28' E, 5 m above sea level) on a Wakanui silt loam (Mottled Immature Pallic Soil, New Zealand soil classification Hewitt, 2010; Eutric Siltic Cambisol, IUSS Working Group WRB, 2015). Soil samples (0–7.5 cm) were taken from triplicated treatment plots (28 × 9 m) consisting of a sheep-grazed permanent pasture (ryegrass (*Lolium perenne* L.) and white clover (*Trifolium repens* L.)), a no-tillage crop rotation (barley (*Hordeum vulgare* L.), wheat (*Triticum aestivum* L.), pea (*Pisum sativum* L.) and ryegrass seed crops) and a permanent fallow treatment (maintained plant-free using glyphosate), 13 years after the trial establishment. Prior to the trial, the site had been under sheep-grazed, ryegrass/white clover pasture for at least 14 years and had a moderately high organic matter content. Further information on soil properties, experimental design and field management has been given by Fraser et al. (2013) and in Subsection 4.1.2 of Chapter 4.

Composites of eight sample cores from each plot were sieved to 4 mm in field moist condition, with visible plant residues and fauna removed, and air-dried before further extraction and analyses. Details on cold (CW) and hot water (HW) extractions, anion exchange resin treatment, and incubation experiment are described in

Subsection 4.1.2. Briefly, CWEOM and HWEOM were sequentially extracted in triplicate with ultrapure MilliQ water (Millipore,  $18.2 \text{ M}\Omega \text{ cm}^{-1}$ ). The soils were extracted at a soil:water ratio (dry weight basis) of 1:10, then centrifuged at  $2831 g$  for 5 min and the supernatants decanted and re-centrifuged at  $17696 g$  for 5 min before filtering through  $0.45\text{-}\mu\text{m}$  polyvinylidene fluoride (PVDF) filters to recover the CWEOM. The residual soil was then re-suspended at the same soil:water ratio and extracted in a water bath at  $80^\circ\text{C}$  for 16 h before repeating the centrifugation and filtering steps to recover the HWEOM. Each CW and HW extract was then treated with an anion exchange resin (Lanxess Lewatit S 5528) at a resin:extract ratio of 1:20 for either 2 h (CW) or 61 h (HW) using a batch equilibrium method.

### 5.2.2 WEOM biodegradability assay

Controlled incubation of the CW and HW extracts, and their respective post-resin treated extracts, was started within 24 h of their extraction and resin treatment. Briefly, 40 mL of each extract were incubated in air-tight, acid-washed, 100 mL Schott bottles at  $25^\circ\text{C}$  for 14 days following the addition of a N-free nutrient solution (2 mL) and 100  $\mu\text{L}$  of fresh soil inoculum. Blanks (MilliQ water), positive (glucose at 20 and 80  $\text{mg C L}^{-1}$ ) and negative (mixed phenol solution comprising quercetin, coumaric, syringic, cinnamic, nicotinic and salicylic acids) controls were run in triplicate with each batch incubation. Headspace  $\text{CO}_2$  concentrations were measured at days 3, 8 and 14 of the incubation, with fluxes integrated over time to assess WEOC mineralization ( $\text{WEOC}_{\text{min}}$ ) expressed as cumulative  $\text{CO}_2$  evolved per unit of WEOC present in the extract at the start of the incubation (Baldock and Broos, 2012). After the incubation, samples were filtered ( $0.45\text{-}\mu\text{m}$  PVDF syringe filters) for further analyses.

### 5.2.3 WEOM characterization

The WEOC concentration was assessed as non-purgeable organic C (NPOC) with a TOC- $V_{\text{CSH}}$  analyzer (Shimadzu Corporation, Kyoto, Japan). The WEON was determined by subtracting the mineral N (Min N) forms ( $\text{NO}_3^-$  and  $\text{NH}_4^+$ ), quantified colorimetrically on a Lachat QuickChem 8500 Series 2 Flow Injection Analysis System (Lachat Instruments, Loveland, CO, USA), from the total dissolved nitrogen (TDN), measured using an on-line UV-catalyzed persulfate oxidation method (Cabrera and Beare, 1993). All extracts were analyzed pre- and post-incubation. Accordingly, the amount of WEON mineralized was determined as the difference in the concentration of Min N forms pre- and post-incubation. The  $\text{SUVA}_{254}$  (specific absorption at 254 nm; Weishaar et al., 2003) and  $a_{254}$  (absorption coefficient at 254 nm; Green and Blough, 1994) were obtained from UV measurements on non- (10-fold diluted) and post-resin treated samples determined on a UV-Vis SpectraMax

190 microplate reader (Molecular Devices Corp., San Jose, CA, USA). Phenolic C concentration was determined with the Folin-Ciocalteu's reagent (Ohno and First, 1998) and expressed as salicylic acid equivalents per unit of WEOC.

#### 5.2.4 Suwannee River humic acid and peat extracts

Extracts of a Suwannee River humic acid (HA) reference standard (IHSS) and a commercially available peat were used as references for the LC-MS analysis. The freeze-dried Suwannee HA standard was dissolved with a 0.5 M NaOH solution. Peat extracts were prepared using either a CW or a 0.1 M NaOH + 0.1 M  $\text{Na}_4\text{P}_2\text{O}_7$  extraction at a sample:extractant ratio of 1:10. Samples were centrifuged and filtered at 0.45- $\mu\text{m}$  (PVDF), as previously described for the CW extracts.

#### 5.2.5 Fluorescence analyses

Fluorescence EEMs were obtained using a FluoroMax<sup>®</sup> spectrofluorometer equipped with a double grating monochromator (Horiba Aqualog Jobin Yvon, France) and a 1.0 cm path length quartz cuvette. Emissions were recorded over a range of 245–828 nm (mean slit bandwidth = 4.66 nm, sum of slit widths = 10), for a range of excitation wavelengths 240–600 nm at 3 nm increments and an integration time of 0.2 seconds. All spectra were normalized for the Raman spectra of ultrapure MilliQ water to eliminate Raman scatter peaks. Intensities were finally expressed in Raman Units (R.U.;  $\text{nm}^{-1}$ ). Rayleigh scatter lines and noise from areas of equivalent emission and excitation wavelengths were removed. The non-resin treated samples were 10-fold diluted to set the WEOC working concentration  $\leq 35 \text{ mg L}^{-1}$  and avoid signal saturation, inner-filter effects and quenching. The resulting EEMs of these samples were finally adjusted for the dilution factor to determine the intensities of the original, undiluted samples. The post-resin treated samples were analyzed in their original form. Peak position based on excitation and emission maxima were assessed according to literature (Coble, 1996; Cory and McKnight, 2005; Fellman et al., 2010).

#### 5.2.6 Preliminary parallel factor analysis

Parallel factor analysis (PARAFAC) was used for dimensionality reduction of the EEMs to a 1-D vector in order to identify statistically independent WEOM fluorescing components (fluorophores) characterized by their unique combination of excitation and emission (Stedmon et al., 2003).

PARAFAC was run on a complete data set ( $n = 216$ ) utilizing an algorithm implemented in the Java programming language at the University of Waikato (<https://github.com/waikato-datamining/multiway-algorithms>). The algorithm was



oriented towards the traditional alternate least square regression approach, as described by Bro (1997). Briefly, PARAFAC initialization was done with 100 iterations and singular value decomposition (SVD) to estimate matrices B and C. Further, matrix A was calculated from the input array matrix X and SVD estimated B and C using the equation:

$$A = X \times ((C| \otimes |B)^{-1})^T \quad (5.1)$$

where  $\otimes$  returns the Khatri-Rao (column-wise Kronecker) product and  $T$  indicates matrix transposition (Burdick, 1995; Bro, 1997).

The same approach was used to calculate B and C, given A and estimated C, or calculated A and B, respectively, and repeated until convergence (lowest reconstruction error, e.g.  $10^{-10}$ ).

### 5.2.7 Data mining

A data mining approach was applied to investigate the relationship between WEOM fluorophores (PARAFAC components), UV and biochemical properties ( $SUVA_{254}$ ,  $a_{254}$  and phenolic C concentration) and biodegradability ( $WEOC_{min}$ , Min N). Data mining of PARAFAC components was performed through the Advanced Data mining And Machine learning System (ADAMS), a modular open-source Java framework (Reutemann and Vanschoren, 2012; Reutemann and Holmes, 2015) supplying a collection of data mining applications, e.g. WEKA 3.9.0 (Bouckaert et al., 2010; Frank et al., 2016), with integrated machine learning methods. Three supervised machine learning methods were tested to evaluate the predictive power of PARAFAC components from the generated models.

Multiple linear ridge regression: a class of linear regression for prediction using the Akaike information criterion for model selection (Akaike, 1974), able to deal with weighted instances. The ridge, defined by  $\alpha$ , was set to 0.00000001 and no attributes were selected, nor were collinear attributes eliminated. This algorithm was mainly used as a baseline to determine the model's lower limit.

Locally weighted-Gaussian Processes (GPD): implemented Gaussian Processes for regression without hyperparameter-tuning, with an inline radial basic function (RBF) kernel (MacKay, 1998). Locally weighted learning (LWL, Atkeson et al., 1997; Frank et al., 2003) was applied using an instance-based algorithm to assign weights to input instances. A neighborhood of 50 similar data rows (on 216 samples) for LWL was determined to build the GPD model for prediction, keeping the default values for data normalization, noise (0.01) and gamma (0.01).

Random Forest (RF): constructs a forest of independent decision trees from randomized subsets of the original variables used to split a node of the tree (Breiman, 2001). In total, 250 trees were constructed. The sum of the predicted numeric values was then averaged across the size of the forest.

Prior to model generation, noisy EEM areas below 290 nm excitation (see Fig. C.2 and C.3) were filtered out. The models were further refined removing mainly two identified outliers. Analytical blanks and controls were left out from the model generation. No pre-processing outside the algorithms was applied to the data generated by PARAFAC. PARAFAC components were then added sequentially to the algorithms to describe the subspace of observed features.

Summary statistics were generated using a 10-fold cross-validation and evaluated for the selection of the best performing algorithm.

### 5.2.8 LC–MS analysis

Analytical triplicates (500  $\mu\text{L}$ ) of non- and post-resin treated extracts were bulked and vortexed before a 400  $\mu\text{L}$  aliquot was filtered with a 0.2- $\mu\text{m}$  PVDF Single Step<sup>®</sup> vial (Thomson<sup>™</sup> P/N: 65531–200). Reference extracts were diluted to 10 ppm with ultrapure water and 400  $\mu\text{L}$  were filtered at 0.2- $\mu\text{m}$  (PVDF) for the LC–MS analysis.

The LC–MS analysis was performed on a Thermo Fisher Scientific<sup>™</sup> (San Jose, CA, USA) Q Exactive<sup>™</sup> Plus Orbitrap mass spectrometer coupled with a Vanquish<sup>™</sup> UHPLC system (Binary Pump H, Split Sampler HT, Dual Oven), controlled by the Standard Instrument Integration for Xcalibur<sup>™</sup> 1.4. (Thermo Scientific, USA). The mass spectrometer was calibrated immediately prior to sample batch analysis with Thermo<sup>™</sup> premixed solutions (Pierce<sup>™</sup> LTQ ESI Positive and Negative Ion Calibration Solutions, Catalog No: 88322 and 88324, respectively). Standards containing caffeine, catechin and trigonelline (1, 10 and 50 ppm) and blank (methanol/water, 1:1 ratio) checks were run before and at the end of each batch. The standard mix (50 ppm) and the sample batch ‘multi-mix’ (equal aliquot of each sample mixed) were run after every 15 sample injections.

The LC was applied to both a hydrophilic interaction (Hypersil Gold<sup>™</sup> HILIC 1.9  $\mu\text{m}$ , 100 mm  $\times$  2.1 mm, P/N: 26502–102130) and a reversed-phase (Accucore Vanquish<sup>™</sup> C18 1.5  $\mu\text{m}$ , 100 mm  $\times$  2.1 mm, P/N: 27101–102130, Thermo Fisher Scientific) liquid chromatography column, hereafter referred to as HILIC and C18 chromatographies, respectively.

A 1  $\mu\text{L}$  aliquot of each prepared extract was separated with either: HILIC conditions – a mobile phase consisting of 0.1% formic acid in acetonitrile (A) and 5 mM ammonium acetate in ultrapure (type 1) water (B), maintained at 55°C and a flow rate of 400  $\mu\text{L min}^{-1}$ , with a gradient elution applied as follows: 0–1 min/5% B, linear increase to 7 min/80% B, isocratic 10 min/80% B, equilibration 11–14 min/5% B; or C18 conditions – a mobile phase consisting of 0.1% formic acid in ultrapure (type 1) water (A) and 0.1% formic acid in acetonitrile (B), maintained at 40°C and a flow rate of 400  $\mu\text{L min}^{-1}$ , with a gradient elution applied: 0–1 min/5% B, linear increase to 7 min/95% B, isocratic to 10 min/95% B, equilibration 11–14 min/5% B.

For both chromatographies, the eluent was scanned from 0.5–11 minutes following heated electrospray ionisation (HESI) at 350°C in both the negative (n) and positive (p) modes with capillary temperature of 320°C. Data were acquired for precursor masses from 110–1600 m/z (HILIC) and 110–1200 m/z (C18) at 70K resolution (AGC target  $3 \times 10^6$ , maximum IT 100 ms, profile mode) with data dependent MS/MS for product ions generated by normalized collision energy (NCE: 35, 45, 60) at 17.5K resolution (TopN 10, AGC target  $2 \times 10^5$ , maximum IT 50 ms, isolation 1.4 m/z).

### 5.2.9 LC–MS data processing

Acquired spectra were processed with the Compound Discoverer v. 3.0 software (Thermo Fisher Scientific). The mass tolerance for spectral selection was restricted to  $\leq 5$  ppm. Only mass spectral peaks for the monoisotopic ion with more than one isotopologue were considered. Compound detection considered multiple ion definitions (details in Appendix C, Section “Processing node 9: Detect Compounds”, “Ions” and “Preferred Ions” of the processing workflow) with a minimum mass peak intensity of 500 000 (Q Exactive Plus). Resolved compounds with a signal-to-noise ratio (S/N)  $> 3$  were assigned elemental compositions using the atom constraints: C<sub>1–90</sub> H<sub>1–100</sub> N<sub>0–15</sub> O<sub>0–20</sub> P<sub>0–1</sub> S<sub>0–5</sub>, an H/C ratio of 0.1–4, and a maximal RDBE of 40, within 5 ppm mass error and intensity tolerance for the isotope pattern search of 30%.

Compound features generated from m/z ions, formulae and spectra features (isotope ratios, precursor and product fragment ions) were queried against KEGG<sup>™</sup>, BioCyc<sup>™</sup>, and Metabolika<sup>™</sup> (Thermo Scientific) biological pathway databases, spectral libraries (mzVault (Thermo Scientific), lipid mediators and polar metabolites libraries (stepped NCE: 10 30 45; Bamba lab 34 and 598, respectively)), ChemSpider and mass search lists (Extractables and Leachables HRAM Compound Database (Thermo Scientific), Biological Magnetic Resonance Bank database (BMRB.db ([www.bmrwisc.edu](http://www.bmrwisc.edu), Ulrich et al., 2008), namely “OpenVK”, sourced from *OpenVanKrevelen*’s repository, Brockman et al., 2018), an in-house library in development (namely “TV”, 55 entries)) for compound identification (see processing workflow in Appendix C). The LC–MS library (97 018 spectra) from the MoNA database (MassBank of North America, <http://mona.fiehnlab.ucdavis.edu/>) was additionally consulted for the identification of masses resolved with the HILIC positive mode.

Differential analyses were applied to sample groups in order to filter compounds of interest for each chromatography/ionization mode. For any given compound, statistically significant differences between sample groups were tested using analysis of variance (ANOVA) followed by a Tukey’s HSD *post-hoc* test at 5% level of significance. Computed *P*-values were then corrected for the whole set of samples (multiple comparisons) as a function of the sample set size (adjusted *P*-value) us-

ing the Benjamini-Hochberg algorithm for the false discovery rate (Benjamini and Hochberg, 1995). Only compounds with adjusted  $P$ -values  $<0.05$  for at least one sample within the whole set were exported for further analysis.

Further, non- and post-resin treated samples and references were corrected for the masses detected in the blank (ultrapure water) and resin controls (resin-treated ultrapure water), respectively, by subtracting the maximum area of a signal in the blank/resin control from the correspondent signal in the sample. For each chromatography/ionization mode batch, median group coefficients of variation (CV) were manually checked and a cutoff of 35% was arbitrarily set in order to account for biological and technical variation (Gan et al., 2007; Casado and Cutillas, 2011; Maes et al., 2015). Data cleaning and all subsequent handling were done in the R environment (R Core Team, 2018).

Out of 2555 identified compounds, only masses with an assigned molecular formula were considered for the data interpretation ( $n = 1708$  for the combined data of both chromatographies and ionization modes). Yet, some significant features whose formula remained unidentified were manually checked with their identification interpreted by automated searches (with reference to theoretical spectra features or literature from SciFinder).

In addition to the constraints applied by the software, heteroatom to C ratios were filtered to  $O/C \leq 2$ ,  $N/C \leq 4$ ,  $P/C \leq 0.5$  and  $S/C \leq 0.8$ , allowing for the extended range (99.99%) of small organic molecules (Kind and Fiehn, 2007). This process filtered further 147 compounds. It should be noted that multiple element count restrictions (Kind and Fiehn, 2007) were verified, but were not implemented in the filtering process given the veiled nature of WEOM and the recognition of naturally occurring metabolites that do not obey this criteria (Pluskal et al., 2012).

Ultimately, for each chromatography and ionization mode, compounds with peak intensity areas below their median value were filtered out (149 compounds).

In R, the unsaturation degree (or double bond equivalent, DBE) of all resolved molecules in WEOM was calculated using the following equation (Koch and Dittmar, 2006):

$$DBE = 1 + \frac{1}{2}(2C - H + N + P) \quad (5.2)$$

and normalized for the number of C atoms (DBE/C) or adjusted by removing the O atom contribution (DBE-O).

The modified aromaticity index ( $AI_{mod}$ ), corrected *a posteriori* by Koch and Dittmar (2016), was defined as:

$$AI_{mod} = \frac{1 + C - \frac{1}{2}O - S - \frac{1}{2}(N + P + H)}{C - \frac{1}{2}O - N - S - P} \quad (5.3)$$

considering that  $1/2$  O is bound through  $\pi$ -bonds (Koch and Dittmar, 2006; Chassé et al., 2015). It is noteworthy that the results ensuing the corrections applied to this formula may not be comparable to previous studies in absolute terms.

Compounds resolved across both chromatographic separations and ionization modes were then classified into six main biochemical categories, namely “Amino sugars”, “Carbohydrates”, “Lipids”, “Nucleotides”, “Peptides” and “Phytochemical oxyaromatics”, defined by a set of constraints, based on C/H/O/N/P stoichiometric ratios, atom number and molecular weight, developed by Rivas-Ubach et al. (2018). If duplicate molecular features occurred due to their detection by both HILIC and C18 chromatographic separations, only the compound with the largest peak area was included for sample classification.

Preliminary results revealed that a relatively large proportion of compounds (34.5%) did not match any of the six classes defined (hereafter referred to as “Not matched”). In order to gain a better characterization of this fraction, assigned formulae were parsed into six discrete categories (Seidel et al., 2014; Chassé et al., 2015), based on their  $AI_{mod}$  (Koch and Dittmar, 2006), H/C and O/C ratios, and heteroatom contents (Perdue, 1984; Stenson et al., 2003): (i) condensed aromatics ( $AI_{mod} > 0.66$ ), such as polycyclic aromatics; (ii) highly aromatic molecules ( $0.66 \geq AI_{mod} > 0.50$ ), including polyphenols and polycyclic aromatics with aliphatic chains (Koch and Dittmar, 2006); (iii) highly unsaturated molecules ( $AI_{mod} \leq 0.50$ ,  $H/C < 1.5$ ), such as lignin-derived material (Stenson et al., 2003); (iv) unsaturated aliphatic molecules without N ( $2 > H/C \geq 1.5$ ,  $N = 0$ ); (v) unsaturated aliphatic molecules with N ( $2 > H/C \geq 1.5$ ,  $N > 0$ ); and (vi) saturated molecules ( $H/C \geq 2$  or  $O/C \geq 0.9$ ).

### 5.2.10 PCA

Principal Component Analysis (PCA) was performed on the quantitative variables and optical indices, fluorescence-derived PARAFAC components and a set of weighted molecular determinants inferred by LC–MS for each combination of land use, extracted fraction and resin treatment, to investigate their relationship with WEOC and WEON mineralization. The nine variables selected from the LC–MS analysis (H/C, O/C, #C, #N, #P, MW,  $AI_{mod}$ , DBE/C, DBE-O) were included in the PCA analysis as a weighted average of each molecular entry (Schmidt et al., 2009):

$$Mv_w = \frac{\sum (Mv_i \times A_i)}{\sum A_i} \quad (5.4)$$

where  $w$  stands for weighted variable,  $Mv$  represents the molecule determinant variable and  $A$  the peak area for each individual molecular formula ( $i$ ). These calculations and PCA analyses were performed in R version 3.5.1.

## 5.3 Results

### 5.3.1 WEOM fingerprinting

For all land uses and both CW or HW non-resin treated extracts, the EEMs revealed a large peak in the area corresponding to the C-peak (Coble, 1996)(see Fig. C.2 and C.3 in the Appendix C to this Chapter). Higher intensities were recorded for the HW extracts, where maxima ranged 1.4–3.2 R.U., with respect to 0.4–1.4 R.U. recorded in the CW extracts (Table 5.1).

A marked decrease of fluorescence intensity occurred following the resin treatment, with values ranging 0.03–0.07 R.U. and 0.16–0.55 R.U. in post-resin treated CW and HW extracts, respectively. The relative decrease of intensity was accompanied by a shift of the emission maximum towards shorter wavelengths (intermediate C-M peak), particularly in the HW extracts (Table 5.1). While the excitation wavelength remained essentially stable for the HW extracts, there was a relative increase for the post-resin treated CW extracts (Table 5.1).

Biodegradation of WEOM from all land uses and both extractions led to a decrease of fluorescence intensity, but the peak position remained relatively unaltered (Fig. C.2 and C.3). In contrast, the magnitude of intensity did not markedly change following incubation of the post-resin treated extracts, but there was a shift of both excitation and emission maxima towards shorter wavelengths, especially in the CW extracts (Table 5.1). Differences between the post-incubation patterns of non- and post-resin treated fluorescence intensities were, however, linked by an increase in WEOC-normalized intensities (Fig. C.4 and C.5).

### 5.3.2 Algorithm evaluation and selection of PARAFAC components

The attributes of the PARAFAC components were investigated through their relationship with WEOM properties and biodegradability. For this purpose, three supervised machine learning algorithms were applied. For each variable and component configuration, evaluations of  $R^2$  and RMSE indicated a greater performance of both RF and LWL-GPD, with respect to ridge regression, in describing WEOM properties and detecting changes in the variables measured.

For all investigated variables other than the phenolic content, the accuracy of the algorithms significantly improved with two components (see Fig. C.8 and C.10 in Appendix C). While the RMSE for Min N remained relatively stable when progressively adding new components to the RF algorithm, it further declined for the other variables while adding up to five (e.g. WEOC<sub>min</sub>) or six (e.g. SUVA<sub>254</sub>) components. With further component configurations, the RMSE varied inconsistently.

**Table 5.1:** Range of maximum fluorescence intensities ( $I_{max}$ ) and relative excitation (Ex) and emission (Em) wavelengths of non- and post-resin treated, cold (CW) and hot water (HW) extracts across three land uses ( $n = 9$  for each land use), before and after the 14-day incubation.

Incubation	Treatment	Extraction						
		CW			HW			
			I <sub>max</sub>	Ex	Em	I <sub>max</sub>	Ex	Em
Pre-	non-resin treated	Min	0.42	330	439	1.44	330	434
		Max	1.45	342	453	3.24	336	443
		Mean	0.66	334	443	2.27	332	440
	post-resin treated	Min	0.03	336	411	0.16	330	406
		Max	0.07	351	443	0.55	339	411
		Mean	0.05	342	430	0.31	336	409
Post-	non-resin treated	Min	0.34	330	434	1.16	333	434
		Max	0.77	339	453	2.48	336	443
		Mean	0.52	335	444	1.77	334	440
	post-resin treated	Min	0.03	324	388	0.12	327	393
		Max	0.05	333	416	0.33	336	411
		Mean	0.04	328	402	0.22	332	405

A relatively similar pattern was observed adding from two up to six or seven components to the LWL-GPD algorithm, albeit, with respect to RF, the RMSE for Min N clearly declined using five components.

The  $R^2$  for the relationship between the observed values and the PARAFAC features generated by the algorithms increased consistently with the decrease in RMSE. Two components explained 93% of the total variation in Min N using the RF algorithm, whereas 70% ( $WEOC_{min}$ ) to 98% ( $a_{254}$ ) were explained by five or six components. The LWL-GPD inferred relatively similar variation patterns across the variables of interest, explaining 95% of the variation in Min N with five components and 73% ( $WEOC_{min}$ ) to 99% ( $a_{254}$ ) using six or seven components for the other variables.

When values for undetected phenols in the post-resin treated fraction were set to zero, as would be their biochemical meaning, none of the algorithms succeeded at establishing a relationship with the pattern observed for the phenols (data not shown). However, when the values of these samples were left out, 74% and 68% of the variation was explained using one component for RF and LWL-GPD, respec-

tively. The  $R^2$  increased up to 86% and 82% by adding six components to the RF and LWL-GPD, respectively, while further additions varied the  $R^2$  inconsistently. Large changes in the  $R^2$  were accompanied with overall small changes in the RMSE when sequentially adding up to six components (Fig. C.9 and C.11), indicating model robustness.

Considering that relatively similar performances were achieved with RF and LWL-GPD, but a relatively faster and less intensive computational process lies behind the RF with respect to the LWL-GPD algorithm, the RF added with six PARAFAC components was elected according to parsimony principles (Table 5.2). Thus, PARAFAC scores of a six-component model were further investigated.

The correlation coefficients between actual and estimated variable patterns are presented in Table 5.2. For all variables, the patterns inferred by six PARAFAC components strongly correlated to actual values.

**Table 5.2:** Summary statistics for the Random Forest regression between actual and estimated variables using six PARAFAC components.

	WEOC <sub>min</sub>	Min N	SUVA <sub>254</sub>	a <sub>254</sub>	Phenols
Correlation coefficient	0.83	0.98	0.96	0.99	0.93
Bias	0.917	0.173	-0.016	1.307	-0.0001
Mean squared logarithmic error	0.05	NaN	0.54	2.44	NaN
Ratio of performance to deviation	1.58	4.36	3.41	7.04	2.55
$R^2$	0.68	0.95	0.93	0.98	0.86
Standard deviation of residuals	50.67	5.07	0.36	32.19	0.12
Mean absolute error	35.11	3.00	0.22	17.98	0.08
Root mean squared error	50.68	5.08	0.36	32.21	0.12
Relative absolute error (%)	50.84	18.34	19.09	10.22	31.10
Root relative squared error (%)	56.56	22.23	27.03	13.85	36.55
Total number of instances	106	106	214	214	106

### 5.3.3 Preliminary filtering evaluation

The median values of CV for group peak areas varied from 13% ( $n = 115$ ) to 22.5% ( $n = 1602$ ) (representing, respectively, the batches with the smallest and the largest number of resolved compounds, Table 5.3), thus indicating relatively high peak reproducibility when all ion intensities were considered (Molloy et al.,



2005). Indeed, a cutoff  $CV \leq 35\%$  allowing for biological and technical (extraction and instrumental) variability removed only  $\sim 0.54\%$  of the total number of resolved compounds ( $n = 2569$ ), mainly derived from the C18(p) batch. This filtering criteria adjusted the CV across all chromatographies and ionization modes within a range of 10.4–11.2%.

Molecules with a predicted formula accounted for 66.9% of the total number of resolved masses. These formulae were further validated by applying heteroatom to C ratios constraints, which filtered 8.5%. An additional 9.5% of the remaining formulae with peak areas below the median value for each chromatography/ionization mode batch, identified with the noise threshold, were removed. About 93% of these were detected in the C18(p) batch. Finally,  $\sim 55\%$  of the initial number of formulae revealed across all chromatographies and ionization modes were utilized for further data analysis.

**Table 5.3:** Initial median values of coefficients of variation (CV) of all peak areas relative to their chromatography/ionization mode, and peak area distribution of only compounds with assigned molecular formulae, following CV and heteroatom/C filtering. The median peak area value laid around the noise threshold and was therefore selected as a filtering criteria for each batch. The number of compounds prior to each filtering step is included in brackets.

	Median area CV (%)	Peak area		
		Median	Mean	Max
HILIC(p)	17.7 (523)	1470 (432)	144 290	46 786 527
HILIC(n)	15.9 (329)	692 (238)	73 554	56 918 433
C18(p)	22.5 (1602)	923 (828)	78 227	34 431 067
C18(n)	13.0 (115)	851 (63)	16 005	670 511

### 5.3.4 Biochemical profile of resolved compounds

Table 5.4 shows the biochemical profile of WEOM resulting from the full data set that included: land use, extraction, resin treatment and incubation effects. Before the incubation, non-resin treated samples were characterized by two dominant

classes, corresponding to peptides (and/or simple amino acids and amides) and compounds not matching any of the classes defined, regardless of CW or HW extraction (Fig. 5.1). Both fractions were, however, larger in the HW extracts. Lipids and phytochemical oxyaromatic compounds comprised the second most abundant proportion of classified compounds, followed by a minor contribution of amino sugars and carbohydrates, and only two if any nucleodites were detected. Similar patterns were observed for all land uses (Fig. 5.1); however, the fallow soil contained fewer compounds from most classes (excluding amino sugars and carbohydrates) in the CW extracts, and showed a reduction of only peptides and not matched compounds in the HW extracts.

**Table 5.4:** Number of molecules and relative proportion of designated biochemical classes across the full data set. Counts refer to unique observations with the highest intensity area across chromatographic separations and ionization modes.

Class	Molecule count	Relative proportion (%)
Nucleotides	2	0.2
Amino sugars	38	2.9
Carbohydrates	35	2.6
Lipids	247	18.5
Phytochemical oxyaromatics	181	13.6
Peptides	369	27.7
Not matched	460	34.5
Total	1332	100

Surprisingly, patterns and relative contributions of these classes remained quite consistent following the resin treatment of CW extracts, with only a minor decrease of not matched compounds in the no-tillage soil. In contrast, substantial decreases of mainly peptides and not matched compounds occurred in the HW extracts, particularly in the pasture soil. It is noteworthy, however, that despite their lower count, the resin treatment of HW extracts removed most of the amino sugars and carbohydrates, along with some lipids and oxyaromatic compounds (pasture and no-tillage).

There was a remarkable decrease in the peptide content of the CW and HW non-resin treated extracts after 14 days of incubation (Fig. 5.1), particularly in

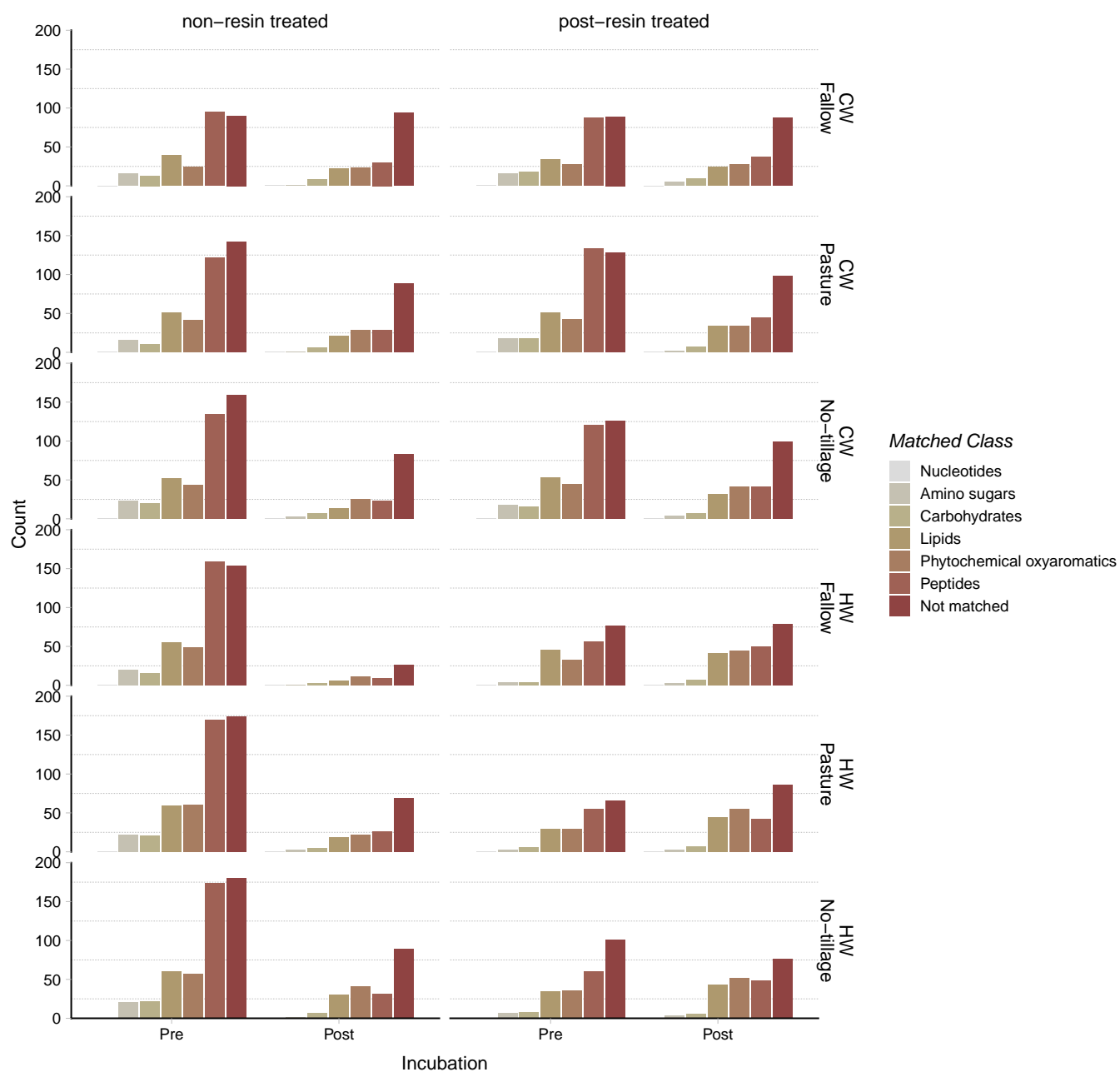
the HW extracts. Such a decrease was observed in the post-resin treated sub-fraction of CW, but not the HW extracts. In the non-resin treated HW extracts of all land uses, the utilization of compounds classified as not matched followed a similar, but less pronounced pattern to that observed for the peptides. Similarly, not matched compounds decreased, to a lower extent, in the non-resin treated CW extracts, with an exception being the fallow soil, where their relative abundance remained unaltered. The same pattern was reflected after the resin treatment of CW extracts, while relatively small if any changes occurred in the post-resin treated HW extracts. While oxyaromatic compounds changed inconsistently with land use in all non-resin treated extracts (with a larger decrease in HW extracts), the other compound classes underwent substantial (lipids), almost entire (amino sugars and carbohydrates) utilization. The reduction of lipids was, however, marginal in the post-resin treated CW extracts, while their proportion remained constant or increased somewhat along with oxyaromatics in the post-resin treated HW extracts. If present, amino sugars and carbohydrates were utilized.

The profile of the not matched category was further investigated by classifying the compounds in six discrete groups based on their  $AI_{mod}$ , H/C and O/C ratios, and N presence. Fig. 5.2 revealed that across all samples, the changes described for the group of not matched compounds were mainly driven by highly unsaturated, possibly lignin-derived moieties (Stenson et al., 2003). Such a fraction prevailed, withal, in the post-resin treated CW sub-fraction, followed by a smaller proportion of aromatic and condensed aromatic molecules with fewer saturated and aliphatic molecules containing N. Overall, the relative contribution of these groups corresponded to that observed in the non-resin treated CW extracts. In contrast, the proportion of highly unsaturated, aromatic and condensed aromatic molecules decreased following the resin treatment of HW extracts, whereas the minor contribution of aliphatic and saturated compounds was not notably affected.

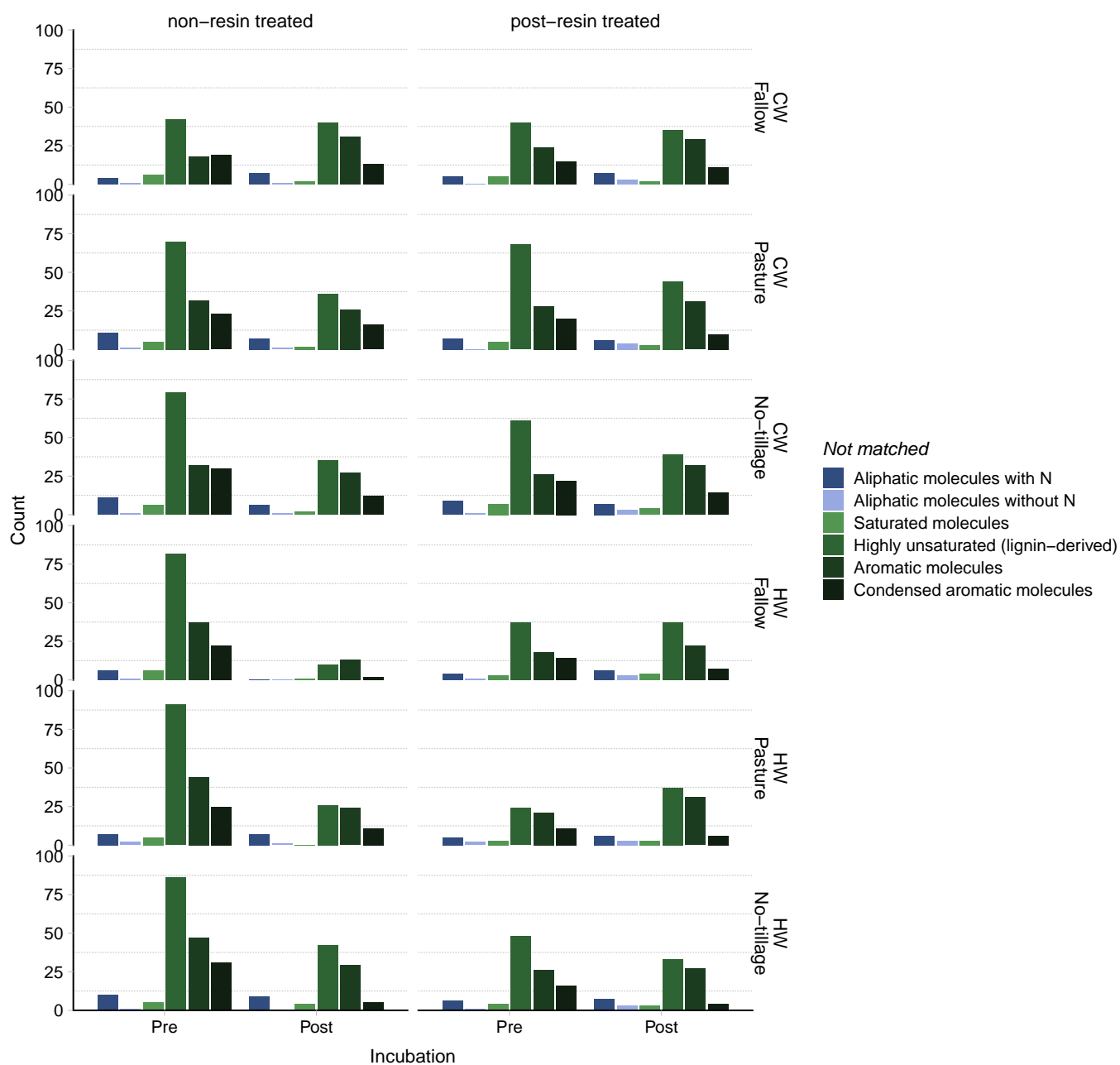
The incubation of non-resin treated HW extracts largely decreased the proportion of highly unsaturated and aromatic moieties, whereas such an effect was less pronounced in the non-resin treated CW extracts, and involved mainly highly unsaturated and condensed aromatics of pasture and no-tillage soils. As previously noted, the same post-incubation trend was observed for the post-resin treated CW extracts, whereas minor or inconsistent post-incubation changes were observed for the post-resin treated HW extracts.

### 5.3.5 Changes in molecular characteristics of WEOM

The van Krevelen diagrams showed the changes occurring in the molecular composition of CWEOM and HWEOM in relation to their N content, DBE-O, DBE/C and MW, after the resin treatment or following biodegradation, inferred by different chromatographies and ionization modes (Fig. 5.3, 5.4, C.16 and C.17). Even



**Figure 5.1:** Distribution of compound classes designated according to Rivas-Ubach et al. (2018) in pre- and post-incubation, non- and post-resin treated, cold (CW) and hot water (HW) extracts of different land uses. Bars represent cumulative counts of unique compounds, regardless of chromatography or ionization mode.



**Figure 5.2:** Distribution of compounds not matching any of the classes defined by Rivas-Ubach et al. (2018) across six discrete categories based on  $AI_{mod}$  (Koch and Dittmar, 2006), H/C and O/C ratios, and heteroatom contents in pre- and post-incubation, non- and post-resin treated, cold (CW) and hot water (HW) extracts of different land uses. Bars represent cumulative counts of unique compounds, regardless of chromatography or ionization mode.

though the C18(p) resolved the largest proportion of compounds across all samples (Table 5.3), the HILIC(p), followed by the HILIC(n), revealed the largest variation due to resin treatment and biodegradation effects within the CW and HW extracts.

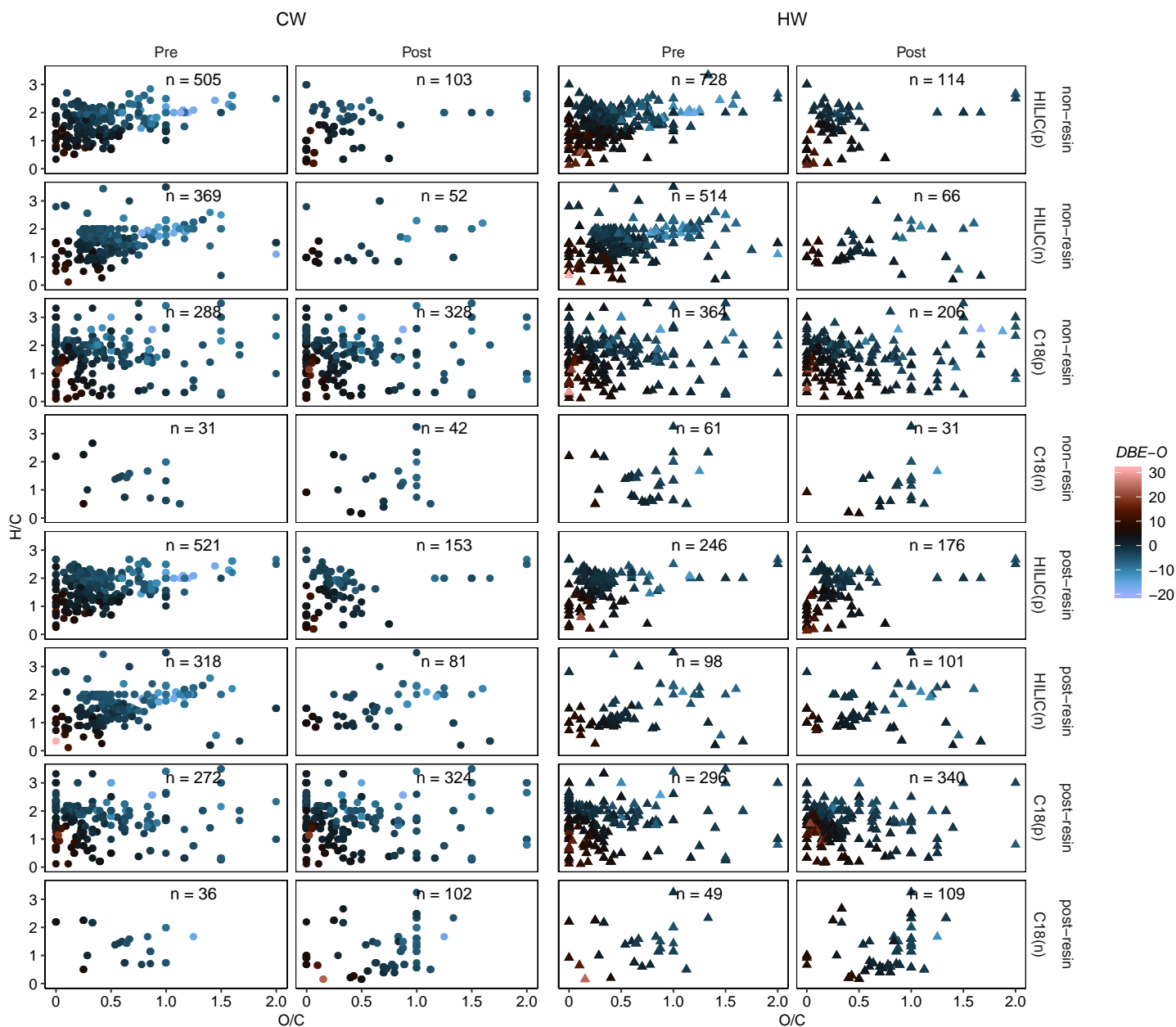
For all extracts, the bulk of the identified compounds had an  $O/C < 1$ , with most values ranging from 0–0.4, and an  $H/C$  ranging from 0–3 (e.g. Fig. 5.3). However, an extended range for  $O/C$  ratios was allowed (Kind and Fiehn, 2007; Pluskal et al., 2012), thus a minor number of compounds displayed  $O/C$  ratios in a range 1–2. There was an increase in the non-resin treated HW extracts of low MW ( $< 500$  Da), N-bearing compounds with an  $O/C < 0.5$  and higher DBE-O with respect to the non-resin treated CW extracts (Fig. C.17, 5.4 and 5.3). Changes in DBE/C were, however, less apparent (Fig. C.16).

Following the resin treatment, the changes revealed by the HILIC chromatography showed a large decrease in the number of molecular formulae in HW but not CW extracts, reflecting the patterns for the peptides and not matched compounds previously observed (Fig. 5.1). As a consequence, post-resin treated HW extracts were characterized by a greater proportion of low MW compounds with higher DBE-O, but with a relatively lower N content and comparable DBE/C with respect to the post-resin treated CW extracts (Fig. 5.3, 5.4, C.16 and C.17). In contrast, relatively smaller changes following the resin treatment were detected by the C18 chromatography.

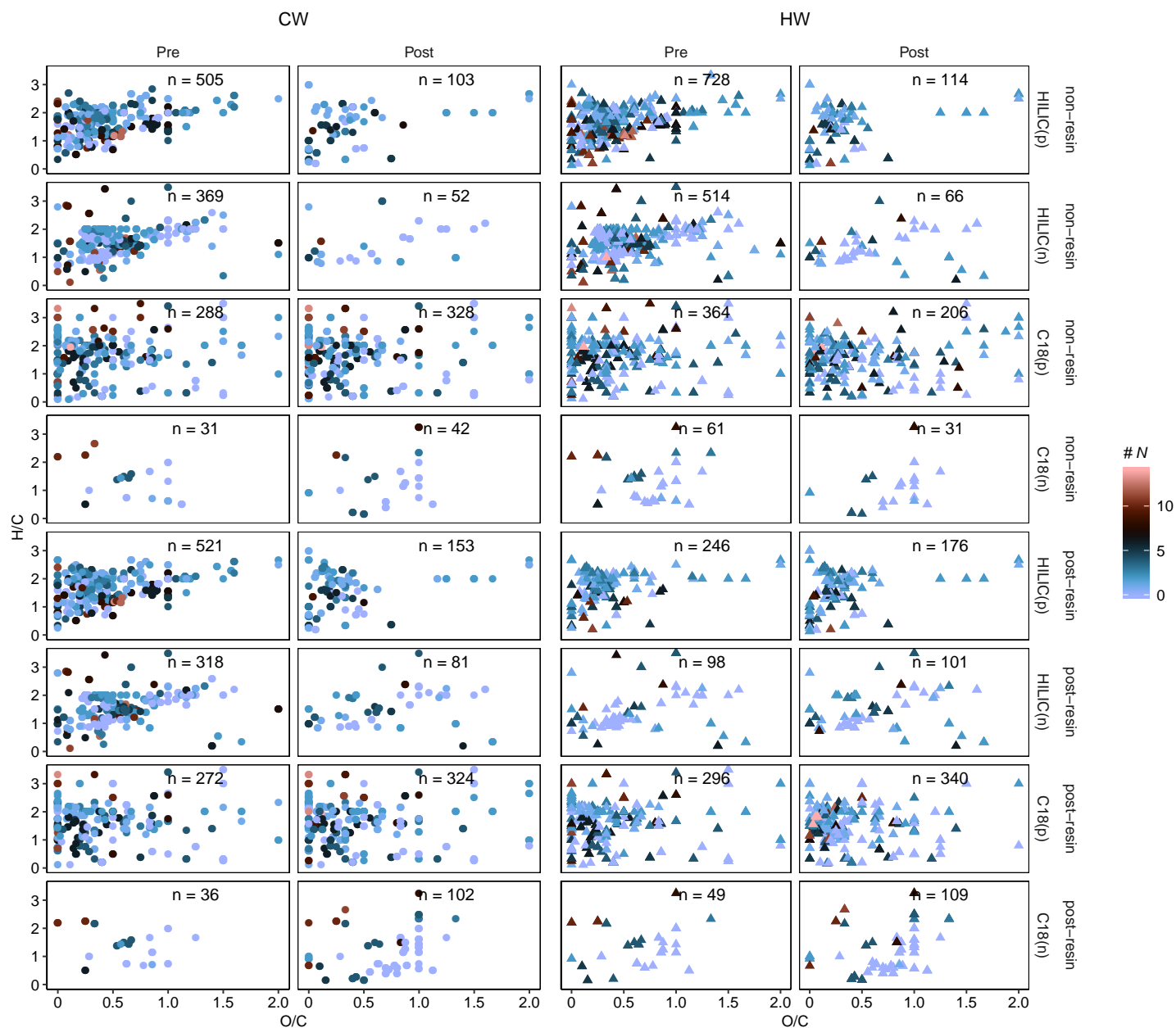
The chromatographic separation revealed a relatively consistent post-incubation trend, regardless of extracted fraction and resin treatment. The HILIC chromatography in both positive and negative modes revealed a net decrease of molecules after 14 days of incubation, with a magnitude of this decrease being larger in the post-resin treated CW extracts and in the non-resin treated HW extracts with respect to their post-resin treated HW and non-resin treated CW counterparts. In contrast, the C18 chromatography inferred a moderate increase in post-incubation compounds, with the exception of the non-resin treated HW extracts. The result of these trends was a net decrease in MW and N content along with a relative increase of the unsaturation degree resolved by the HILIC chromatography, but inconsistent post-incubation variations in MW, N content and unsaturation degree revealed by the C18 chromatography. It is noteworthy that there was an increase in high MW, N-bearing molecules with low unsaturation degree resolved by the C18 in the post-resin treated HW extracts (Fig. C.17).

### 5.3.6 Optical and molecular aspects in relation to WEOM biodegradability

Mean PARAFAC scores and weighted averages of the molecular variables included in the PCA are available in Tables C.2 and C.3 of Appendix C. Generally, the



**Figure 5.3:** van Krevelen diagrams of non- and post-resin treated, cold (CW) and hot water (HW) extracts before and after the 14-day incubation, plotted against DBE-O and relative to their chromatography/ionization mode of detection. n = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes).



**Figure 5.4:** van Krevelen diagrams of non- and post-resin treated, cold (CW) and hot water (HW) extracts before and after the 14-day incubation, plotted against the number of nitrogen atoms (# N) and relative to their chromatography/ionization mode of detection. n = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes).

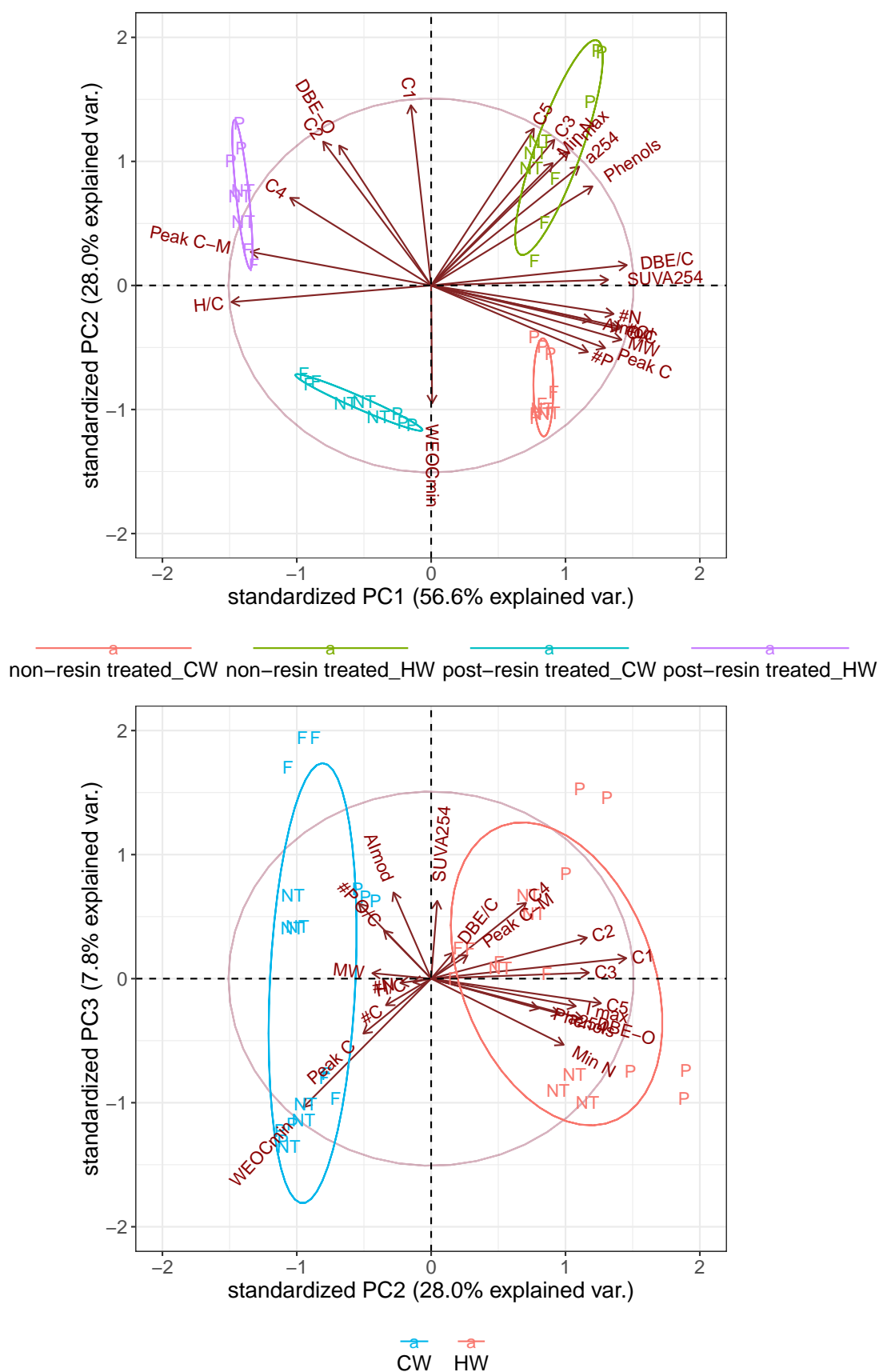


PARAFAC scores of all components were larger in HW extracts than in the CW extracts and consistently increased from fallow to pasture soil extracts. The scores revealed by PARAFAC component 1 were by far the largest in both non- and post-resin treated HW extracts. Component 3, followed by component 5, displayed higher scores in the non-resin treated HW extracts, whereas scores of components 2 and 4 were larger in their post-resin treated counterparts. The scores of component 6, resulting from a six-component PARAFAC model, appeared to be negative. While negative scores do not have any physical meaning, it has been suggested that these might be the result of a simplification of complex systems of natural/biological origin, where a continuum of similar fluorophores attached to distinctive macromolecules is described by discrete PARAFAC modeling (Thygesen et al., 2004). Nevertheless, for the purposes of PCA, and according to non-negativity constraints, this component was not further explored.

Despite relatively small differences in the number of compounds revealed in the non- and post-resin treated CW extracts, there was a general tendency for MW, O/C, DBE/C,  $AI_{mod}$  and the number of O and P atoms to decrease and H/C to increase after the resin treatment of CW extracts (Tables C.2 and C.3 in Appendix C). The impact of a larger number of O atoms in the non-resin treated CW extracts was reflected in a relatively lower DBE-O with respect to the post-resin treated CW extracts. The same pattern was observed after the resin treatment of HW extracts, with an additional mean decrease in the number of N atoms (Tables C.2 and C.3).

These changes were revealed by principal components 1 (PC1) and 2 (PC2), which together explained 84.6% of the variance among the extracts (Fig. 5.5) when including the molecular information from all analyses. PC1 revealed 56.6% of the variance derived from the resin treatment. The PARAFAC components 3 and 5 dominated in the non-resin treated extracts, and particularly in the HW extracts characterized by a relatively larger phenolic content,  $a_{254}$  and Min N. The larger DBE/C and  $SUVA_{254}$  expressed in these samples strongly correlated with PC1. PARAFAC components 2 and 4 were prominent in the post-resin treated, mainly HW extracts, featured by a larger DBE-O. In contrast, component 1 was weakly projected on to PC1. There was an association between the intermediate C-M fluorescence peak and the H/C ratio on PC1. The variation in  $WEOC_{min}$  was totally uncorrelated to these changes, yet was negatively projected on to PC2, which explained further 28% of the total variation. PC2 was characterized by differences in DBE-O,  $a_{254}$  and phenolic C content, which separated CW and HW extracts, regardless of resin treatment (Fig. 5.5). All PARAFAC components were positively projected on to PC2, and in particular component 1.

The variation caused by the land use was discernible to some extent only in the HW extracts and was largely reduced in the CW extracts and respective post-resin treated sub-fractions.



**Figure 5.5:** Principal component analysis of PARAFAC component scores (C1–C5), weighted molecular variables, WEOM optical properties and biodegradability indicators. Axes were rotated to visualize the vector loadings in the second dimension (PC2). P, pasture; F, fallow; NT, no-tillage. Variables' abbreviations are defined in text.

## 5.4 Discussion

### 5.4.1 RF identification of PARAFAC components

EEM spectral decomposition by PARAFAC modeling enabled the identification of distinct fluorescent components capable of describing the variation in WEOM biodegradation and its properties resulting from the interaction between land use, extracted fraction, resin treatment and biodegradation effects.

Ohno and Bro (2006) previously pointed out the relative importance of the sample size in resolving fluorophores with similar spectral properties. PARAFAC ran on a data set comprising of 216 samples and forward search of the optimal number of explanatory components using the RF algorithm revealed six components foremost explaining the variability of all the selected variables, thereby suggesting the strong link between fluorescence features and WEOM biodegradability and biochemical properties (Kalbitz et al., 2003; Saadi et al., 2006; Baker et al., 2008; Fellman et al., 2008). Drifts in the variation explained by a larger number of components suggested that the algorithm was in fact inferring patterns from noise. On the other hand, values of RMSE and  $R^2$  suggested that two large fluorophores were associated with most of the variation in WEOC<sub>min</sub>, Min N and UV properties, confirming the strong effect of the aromatic and/or conjugated character of WEOM on biodegradability.

The best model fit for the phenolic C concentration led to the prediction of negative values. Such a result alludes to the phenolic C content in the post-resin treated sub-fraction, along with that in the post-incubation CW extracts, which was below the detection limits set by the calibration curve for the Folin-Ciocalteu's method. Actual null values would then need to be set as such in the post-processing step of pattern estimation. In contrast to the other variables, the bulk of the changes in the phenolic C content was described by a singular PARAFAC component, confirming the large impact of the phenolic hydroxyl groups on the fluorescence properties of WEOM (Senesi et al., 1991). An increase of up to six components, thus, suggested either differences in polycondensation and/or polysubstitution or different structural features that can explain fluorescence in spite of their paucity.

### 5.4.2 Analysis of EEMs and PARAFAC components

Differences in fluorescence intensities between CW and HW extracts were largely driven by ~3.6-fold higher HWEOC concentrations. According to Coble (1996), the large C-peak identified in the CW and HW extracts indicated the presence of plant and SOM-derived, fulvic-like WEOM. Yet, higher intensities in the HW extracts were mainly associated with the larger phenolic C content (see Table 4.3 in Chapter 4). The C-peak was consistently found in extracts of all land uses,

indicating the presence of the same type of fluorophores, but their contribution to fluorescence intensity in CW and HW extracts varied according to changes in their concentration across different land uses. This variability was more pronounced in HW extracts, suggesting that this fraction is highly dynamic.

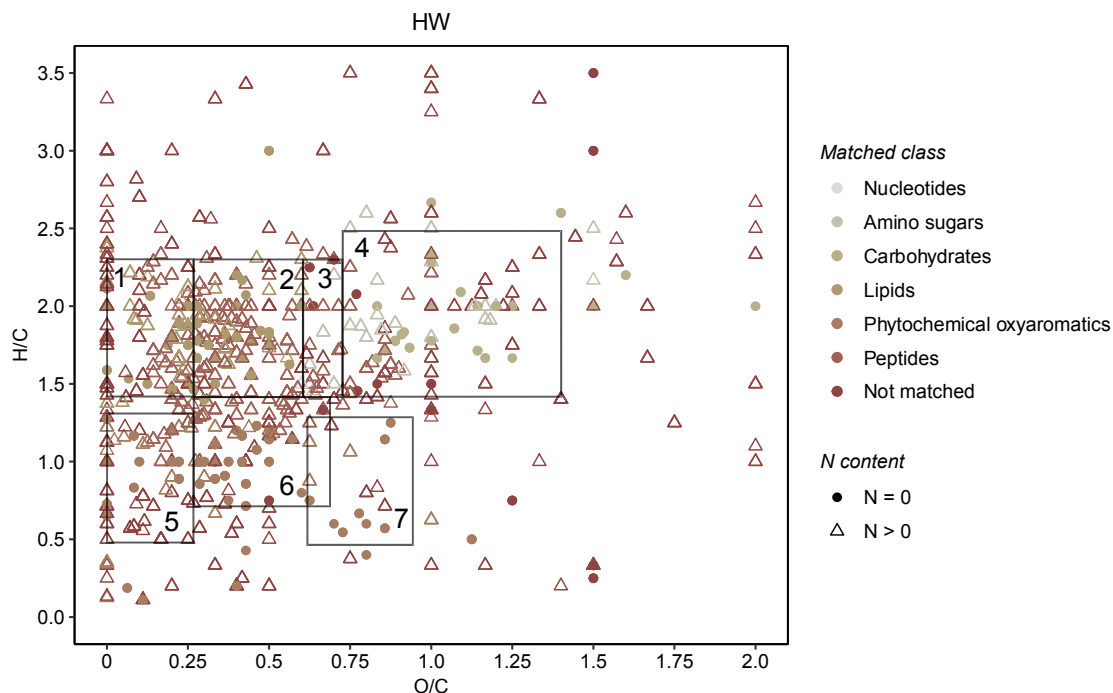
Following the resin treatment, the reduction in fluorescence intensity confirmed that the bulk of the CWEOM and HWEOM comprised of humified material and phenolic moieties derived from the decomposition of plant-derived organic matter (e.g. lignin). In contrast, the remaining post-resin treated WEOM proportion was characterized by blue-shifted emission maxima indicating an intermediate C-M peak consistent with a small fraction of fluorescing by-products of microbial metabolism. This shift was accompanied by a decrease in  $a_{254}$  (see Fig. C.1 in Appendix C, and further evidence in Fig. 5.5), indicating a close relationship between fluorescence and absorbance (Green and Blough, 1994). For instance, Cory and McKnight (2005) observed similar excitation and emission spectra referred to less conjugated forms of reduced quinones of microbial origin or microbially derived moieties surrounding the fluorophore. The latter seems to be a more plausible explanation, considering that a hydrophobic interaction of the anion exchange resin with hydroquinones might have occurred. Yet, it can be speculated upon the involvement of intramolecular electron donor-acceptor complexes formed through charge-assisted H-bonds (Ni and Pignatello, 2018) between e.g. functionalized derivatives of quaternary alkylbenzenes and hydroquinones ensuing microbial metabolism.

The decrease of fluorescence intensity following biodegradation suggested the relative utilization of aromatic and conjugated moieties over a short period (14 days). This was further confirmed by the corresponding decrease in  $a_{254}$ , thus demonstrating a strong ability of PARAFAC components to interpret changes in  $a_{254}$  (Table 5.2). However, WEOC-normalized fluorescence intensities revealed an increase in the relative amount of fluorescing units per unit of WEOC following biodegradation of both non- and post-resin treated extracts (see also Fig. C.4 and C.5 in Appendix C), indicating that both plant and SOM-derived, as well as microbially derived fractions comprised of a significant proportion of unreactive WEOM. It is noteworthy that such an increase cannot exclude the formation of new fluorescing material following biodegradation (Saadi et al., 2006), an effect that might be outweighed by concomitant degradation processes, resulting in a net reduction of the post-incubation fluorescence signal. The latter scenario might involve reduction processes ensuing WEOM biodegradation, thus leading to relatively more intense fluorescence intensities (Cory and McKnight, 2005). The increase in the proportion of fluorescing units following biodegradation was reflected by the relationship between  $SUVA_{254}$  and the component features estimated by the RF algorithm (Table 5.2), denoting fluorescence of aromatic and conjugated systems (Senesi, 1990). Four clusters emerged (Fig. C.12), corresponding to post-incubation non-resin treated

CW (highest values, 4–5.1 L mg<sup>-1</sup> m<sup>-1</sup>) and HW (higher central values, 2.7–3.5 L mg<sup>-1</sup> m<sup>-1</sup>) extracts, pre-incubation non-resin treated CW and HW extracts (lower central values, 1.8–3.1 L mg<sup>-1</sup> m<sup>-1</sup>) and pre- and post-incubation post-resin treated extracts (lower and upper end, respectively). While these patterns cannot discern between the progressive transformation of WEOM towards more aromatic, condensed or conjugated structures (non-resin treated extracts) or their accumulation following the depletion of more labile WEOM, they clearly show a consistent change in the composition of “newly available” substrate. These changes were further investigated using van Krevelen diagrams for WEOM profiling.

### 5.4.3 Current understanding of van Krevelen diagrams

The bulk of the O/C and H/C ratios revealed across all WEOM samples, was characteristic of DOM and WEOM (Kim et al., 2003; Ohno et al., 2010; Roth et al., 2015; Verkh et al., 2018; D’Andrilli et al., 2019). Resolved compounds were most abundant in an area corresponding to lignin derived moieties, proteins (and/or peptides and amino acids) and lipids (Ohno et al., 2010; D’Andrilli et al., 2015; Brockman et al., 2018), hereby identified as peptides, lipids, phytochemical oxyaromatics and not matched compounds (Rivas-Ubach et al., 2018). While there was a large overlap of not matched compounds with all other classes (see Fig. C.18 and C.19 in Appendix C), most of these compounds were centered in the lignin, protein, lipid and condensed aromatic hydrocarbons regions (e.g. Fig. 5.6), hence according to the patterns described in Fig. 5.2. It appeared from Fig. 5.2 that the bulk of the not matched compounds comprised of highly unsaturated, possibly lignin-derived (Stenson et al., 2003), and aromatic moieties. However, the fact that phenolic C, a typical biomarker of lignin decomposition (Hedges and Ertel, 1982; Senesi et al., 1991), prevailed in the HW extracts with respect to the CW extracts (Table 4.3 and supplementary Fig. C.1), suggests that decomposition processes, involving ring opening reactions and progressive oxidation of aliphatic chains of phenolic moieties to carboxyl groups (Waksman, 1932), would have occurred during the formation of the CWEOM, reducing the phenolic C signature without changing the position of these new compounds from the lignin region (Sleighter and Hatcher, 2008). Furthermore, given the ubiquitous presence of N in these compounds, such an effect might have also involved the oxidation of lignins to quinones in the presence of amino compounds (see Stevenson, 1994 for more detail) and/or reactions with N-containing nucleophiles. In support of these hypotheses is a general decrease in DBE-O of the CW extracts, indicating a higher degree of oxygenation of this fraction (Fig. 5.3), referred to as fulvic acid fraction (Stenson et al., 2003). In this context, Perdue and Ritchie (2003) reported the C oxidation state of humic substances to range +0.25–0.33 with respect to their precursors, having an overall C oxidation state of -1. Considering an initial oxidative step, this would correspond



**Figure 5.6:** van Krevelen diagram of all compounds revealed in pre-incubation, non-resin treated hot water extracts across different land uses. Designated areas were approximated from literature and correspond to: 1. lipids, 2. proteins, 3. amino sugars, 4. carbohydrates, 5. condensed hydrocarbons, 6. lignin, 7. tannins. The color code corresponds to the classes defined according to Rivas-Ubach et al. (2018).

to greater O-functionalized unsaturation of the transformed material, without fragmentation of the C backbone. This might explain the relatively narrow window of MW observed for CW and HW extracts.

For both CW and HW extracts, the location of phytochemical oxyaromatics resolved by the HILIC(n) within the van Krevelen plot with O/C 0.2–0.4 and H/C  $\sim$ 0.9 suggests that lignin was their common source. Supporting evidence is given by the presence of well documented lignin derivatives such as 4-hydroxybenzoic acid, 4-hydroxybenzaldehyde, 4-hydroxy-3-methoxycinnamaldehyde, *trans*-cinnamic and *p*-coumaric acids (Waksman, 1932; Whitehead, 1964; Kononova, 1966; Mansell et al., 1974; Boerjan et al., 2003; Dai et al., 2018; Takada et al., 2018), as well as quinones (e.g. 2-methyl-1,4-benzoquinone), all products of lignin degradation and redox reactive components of humic substances (Senesi, 1990; Newman and Kolter, 2000; Cory and McKnight, 2005).

Phytochemical oxyaromatics, along with not matched compounds, also spanned the adjacent region of the van Krevelen diagrams commonly assigned to condensed aromatic hydrocarbons and resolved in the positive ionization mode. This was consistent with the presence of N in their structure (Fig. 5.4, C.18 and C.19), suggesting a potential mechanism for its preservation (Hsu and Hatcher, 2005) or stabilization on soil mineral surfaces (Ohno et al., 2010).

An interesting area populated by not matched compounds in both CW and HW extracts with an  $O/C = 0$  indicated a large proportion of non-oxygenated, low MW compounds varying from highly unsaturated, condensed hydrocarbons ( $H/C \leq 1.5$ ) to fully saturated molecules ( $H/C > 2$ ). Whilst most of these compounds remained unknown, some were identified as (amino) benzothiazoles and dibenzylamines ( $H/C < 1.4$ , Table C.6), reported as either micropollutants with potent antimicrobial activity (Xiao et al., 2016) or bacterially produced fungistatic volatiles (Zou et al., 2007), adenine ( $H/C = 1$ ), or cyclic amines (e.g. 2-ethylpiperidine,  $H/C = 2.1$ , Table C.6) possibly derived from heteroaromatization of aliphatic imines. Clearly, the presence and impact of these and similar compounds on WEOM biodegradation processes is not inconsequential, as partly revealed by the van Krevelen post-incubation plots (Fig. C.18 and C.19), and thus requires more investigation.

Lipids and lipid-like moieties overlapped in the area largely populated by peptides and amino-N. Lipid biomarkers have been suggested to be used as tracers of SOM biogeochemical dynamics due to their selective preservation potential and structural specificity (von Lützow et al., 2007; Filley et al., 2008; Pisani et al., 2014), whereas amino-N compounds are known for their ease of biodegradation and reutilization for biological production. Amino sugars contributed only to a minor extent to the total WEOM pool and included D-glucosamine, DL-2-amino adipic acid, N-acetyl-D-mannosamine, 2-acetamidoglucal, glycerophosphocholine, neuraminate and another 57% of amino sugar-like compounds. Similarly, the abundance of carbohydrates ( $O/C \geq 0.8$  and  $1.65 \geq H/C < 2.7$ , Fig. 5.6) was relatively small, which would have been expected for soils collected in early spring, indicating that they have been depleted during the winter period.

Resin treatment changes revealed by the HILIC chromatography confirmed that the bulk of HWEOM comprises of low MW, relatively unsaturated N-bearing domains and saturated N compounds, alluding to the proportion of not matched, oxyaromatic, and peptide-like compounds (Fig. 5.3, 5.4, C.16, C.17, C.19). However, the C18 chromatography emphasized the relevant contribution of small uncharged or protonated, relatively unsaturated N compounds comprising the HWEOM sub-fraction isolated by the resin treatment. In contrast, the abundance of low MW, N compounds revealed by both chromatographies in the post-resin treated CW sub-fraction suggested that these compounds comprise a larger proportion of the

CWEOM, and that the changes in WEOC concentration following the resin treatment related to only a specific fraction of the whole extract, without significantly altering the composition of the medium. According to fluorescence measurements, such a fraction would refer to the fulvic-like fraction, comprising of highly carboxylated, polymeric units, identifiable within the lignin region. However, reasons for the overall poor exchange of compounds revealed under the HILIC(n) in the CW extracts are not obvious, but one explanation might consider the pH variability among extracts from different land uses with respect to the pKa of these compounds or, for instance, the lack of ionization of hydroxy groups on alicyclic compounds.

Post-incubation changes revealed by the HILIC chromatography suggest the utilization of lignin-derived compounds from all extracts. In fact, the characteristic lignin biomarkers were no longer present in the post-incubation extracts, with exception of 4-hydroxybenzaldehyde, and partly 4-hydroxybenzoic acid, recovered in all fractions of grassland and no-tillage soils. This is in contrast with the opinion regarding lignins as refractory compounds (D'Andrilli et al., 2013) and suggests their dynamic involvement in biodegradation pathways. Indeed, both aerobic and anaerobic degradation of lignin-monomers have been demonstrated (e.g. Fuchs et al., 2011). This suggests that an appreciable proportion of compounds revealed by the C18(p) chromatography in the lignin region, not identifiable as lignin-derived, is either resistant to biodegradation or likely represents newly synthesized by-products of microbial activity (e.g. post-incubation, post-resin treated HW extracts in Fig. C.17). Circumstantial evidence alludes to the effect revealed by fluorescence EEMs, mainly referring to the mean increase in post-incubation fluorescence intensity normalized per unit WEOC (Fig. C.4 and C.5). Accordingly, one would also assume a relationship between small hydrophilic neutrals, positively ionizable WEOM compounds and fluorescing moieties. Earlier studies reported the presence of carboxylic-rich alicyclic molecules (CRAM, Hertkorn et al., 2006) along with nonprotonated carbon centers in aromatic heterocycles and nonprotonated quaternary carbon (Gao et al., 2016a; Cao and Schmidt-Rohr, 2018), associated with substitution, branching, and cross-linking occurring during humification (Mao et al., 2007). Their occurrence has been attributed to, respectively, progressively demethylated, terpenoid-like molecules (Leenheer et al., 2003; Hertkorn et al., 2006), tannins and tannin-like structures (Kögel-Knabner et al., 1991), but also aromatic amino acid residues, such as those of tyrosine and phenylalanine (Glick, 2009), and N-containing alkyl-substituted aromatic heterocycles (Mao et al., 2007). These compounds, which have been conferred recalcitrance (Kögel-Knabner et al., 1991; Hertkorn et al., 2006), overlap with the lignin region of the van Krevelen diagram (Hertkorn et al., 2006; D'Andrilli et al., 2013) and may be, therefore, linked to the proportion of undefined, relatively unsaturated, post-incubation compounds revealed mainly by the C18 chromatography in the current study. Complementary



analyses, such as  $^{13}\text{C}$  solid state and  $^1\text{H}$  NMR, and further investigation are necessary to confirm these ideas and the role of these and simile structures in WEOM biodegradability studies.

Contributing to the pool of low molecular weight, unsaturated compounds present after the incubation were (N-containing) lipids and (condensed) phytochemical oxaromatics, but also peptides and/or protein-like compounds (saturated), which would have contributed to the residual fluorescence intensity in both non- and post-resin treated extracts. Of note is the increase, in the post-resin treated (especially HW) samples, of compounds in the lipid region, which has been also associated with CRAM-like structures (DiDonato et al., 2016).

Although these van Krevelen representations with N-bearing compounds might appear oversimplified, they provide a useful indication of the WEOM forms involved in biodegradability processes. The major classes emerging in this study, according to the multidimensional stoichiometric constraints classification approach defined by Rivas-Ubach et al. (2018), were further investigated.

#### 5.4.4 Peptides or nonpeptides?

The large proportion of compounds classified as peptides revealed in both CW and HW extracts indicates a large contribution of high microbial activity and microbial biomass turnover in the formation of WEOM (Kögel-Knabner, 2006). In the CW extracts, the bulk of these compounds comprised the post-resin treated sub-fraction, supporting the idea of the microbial origin of this sub-fraction (see Chapter 4). However, their low affinity for the anion exchange resin in the CW extracts has been questioned, as for many, their retention due to carboxylic groups would have been expected. A dominant peak corresponding to betaine (Table C.6), a quaternary ammonium compound, occurred in both non- and post-resin treated CW extracts of all land uses. Microorganisms accumulate betaine in response to osmotic stress conditions, such as soil drying (Warren, 2014), thus its large presence in CW extracts might have occurred following cell lysis (Schimel et al., 2007) or release following soil rewetting (Halverson et al., 2000; Warren, 2013a). Its plant origin and release as an osmoprotectant cannot, however, be excluded. Goldmann et al. (1991) have shown that soil bacteria, and in particular *Rhizobium sp.*, are capable of utilizing exogenous betaines as a source of C and N during osmoregulation. However, the mechanisms supporting betaine biodegradation as a readily available WEOM source cannot be fully confirmed by the present study, given that it has been consistently found in the post-incubation extracts of the post-resin treated, but not non-resin treated CW and HW fractions. Further confirmation of its likely microbial derivation is, indeed, given by its presence in both non- and post-resin treated HW extracts, resulting from cell lysis at high extraction temperatures (Sparling et al., 1998).

Other osmolytes, corresponding to ectoine (Table C.6), a pyrimidine derivative (Lippert and Galinski, 1992), and L(-)-carnitine (Table C.6), also a quaternary ammonium compound, were identified within the same class in pre-incubation extracts. While L-carnitine was found in all extracts, ectoine was mainly present in non- and post-resin treated CW extracts of all land uses. In addition to their osmoregulatory function, these compounds have also been identified as C and N sources for bacterial cellular functions in both aerobic and anaerobic environments or as final electron acceptors (Talibart et al., 1994; Meadows and Wargo, 2015). This is consistent with their utilization observed in the current study and likely explains the post-incubation occurrence of betaine in post-resin treated extracts as a result of an uncomplete L-carnitine metabolism (Meadows and Wargo, 2015). Following this pathway, L-carnitine is used as a N source, following serine deamination to pyruvate and ammonia. This is in agreement with the complete mineralization of betaine in non-resin treated CW and HW extracts, and suggests the contribution of quaternary ammonium compounds to the  $WEOC_{min}$  and Min N pools (Fig. 4.1 and B.4). In contrast, ectoine assimilation and utilization follows a different pathway, with  $CO_2$  ultimately released and no ectoine-derived metabolite accumulation (Talibart et al., 1994).

Similarly, crotono-betaine (Table C.6), a betaine-fatty acid derivative, was also not retained by the resin treatment of both CW and HW extracts, likely due to the protonation of its zwitterionic structure at low pH and the presence of quaternary ammonium groups (Table C.6). Such a compound is another known intermediate of the L-carnitine metabolism in microorganisms, further acting as electron acceptor under anaerobic conditions and absence of nitrate (Roth et al., 1994). When lacking other readily available C sources, such as glucose or succinate, bacteria adapted to these conditions (e.g. *E. coli*) utilize betaine derivatives under coordinated induction (Collins et al., 1997) to satisfy their metabolic needs.

Recent studies conducted by Warren (2013a,b) showed that quaternary ammonium compounds in sub-alpine grasslands and wheat seedlings were among the top 10 most abundant N compounds. Nevertheless, their impact on WEOM cycling has been largely overlooked (Warren, 2019). Although their net fluxes were shown to be several magnitudes smaller than rapidly degradable amino acids (e.g. alanine) using laboratory incubations, their cycling in soils exposed to environmental changes, such as fluctuations in soil water content, may be considerably larger (Warren, 2019). The results presented here, thus, warrant their inclusion in WEOM studies and invite further investigation of their impact on WEOM mineralization.

Both CW and HW non-resin treated extracts comprised of L-amino acids (methionine, proline, phenylalanine, valine, asparagine, threonine, tyrosine, aspartic and glutamic acids) and L-enantiomers of small di-peptides (e.g. alanyl-proline, glycyl-proline, glycyl-phenylalanine, alanyl-leucine, isoleucyl-alanine, isoleucyl-

glutamine, isoleucyl-serine, leucyl-valine, leucyl-alanine, leucyl-asparagine, prolyl-isoleucine, seryl-leucine,  $\beta$ -homoalanyl- $\beta$ -homoleucine), generally utilized over the short-term incubation. However, an inconsistent post-incubation permanence of proline (CW), alanyl-proline (HW), glycyl-proline (CW and HW) and seryl-leucine (CW) was noted in the non-resin treated extracts, regardless of land use. For instance, the preservation of proline, a well recognized osmoprotectant (Csonka, 1989; Kempf and Bremer, 1998) of several bacteria (e.g. *Bacillus sp.* and *Rhizobium sp.* under drought stress and salt resistance, Gloux and Le Rudulier, 1989; Chen et al., 2007), might indicate a regulatory mechanism for microbial support in adverse conditions and selective utilization. Further, strong electrostatic (Biedermannova et al., 2008) and hydrophobic (e.g. Zang et al., 2000) interactions with WEOM aromatic moieties favored by its structural features (e.g. heterocyclic structure) or intrinsic functional properties (see Redmile-Gordon et al., 2015 and references therein) cannot be excluded. Interestingly, DL-phenylalanine was identified in only pre-incubation non-resin treated CW extracts and conversely in post-incubation non-resin treated HW extracts. These results suggest that pathways of biosynthesis and biodegradation are driven by the microbial nutrient partitioning in response to the composition of the surrounding media, possibly involving different microbial co-workers. It is noteworthy the L- configuration of nearly (if not) all amino acids and peptides, suggesting their relatively recent formation, given that racemization to their respective D-enantiomers is a slow process resting on both biotic and abiotic processes (Amelung, 2003). A small fraction of D-amino acids in soils also originates from bacterial peptidoglycans, hence their absence in the extracts hereby investigated may be also due to plant uptake at soil solution concentrations (Hill et al., 2011). Although these hypotheses still need to be confirmed (e.g. Amelung, 2003), they align with the idea of a more recent, microbial origin of the post-resin treated sub-fraction (Chapter 4).

Indeed, most of the amino acids and peptides were also found in the post-resin treated CW extracts (with exception of e.g. leucyl- and isoleucyl-alanine), yet only tyrosine and a few peptides were not retained by the resin treatment of HW extracts (e.g. alanyl-proline and some proline- and (iso)leucyl-bearing ends). These general patterns, observed in Fig. 5.1, were likely due to the isoelectric point of amino acids and peptides at the pH of CW and HW extracts of the different land uses (see Table 4.3 in Chapter 4). These few oligopeptides revealed in the post-resin treated HW extracts consistently occurred in both post-incubation CW and HW post-resin treated extracts of at least one land use. Their poor involvement in biodegradability was partly reflected in their non-resin treated counterparts. Isoleucine, for instance, has been recognized as part of fulvic acids in NMR studies conducted by Mao et al. (2007); tyrosine might have been involved due to nonprotonated aromatic structures, as discussed earlier. However, further investigation is needed to

clarify the controls of their utilization in the soil solution. In addition, although the presence of tyrosine has not been clearly observed in the fluorescence EEMs, its involvement in the emission blue shift (C-M peak) cannot be excluded. These findings support the hypothesis that the pathways of soil C and N utilization will be regulated by a suite of low MW peptides actively involved in WEOM biodegradation (Farrell et al., 2013) and is in agreement with recent studies re-appraising the likely underestimated size of this pool (Farrell et al., 2011). More work on the factors affecting the molecular characteristics, hence solubility and concentration of small peptides in soil water is, therefore, essential.

In addition to quaternary ammonium compounds and protein amino acids, non-protein amino acids were involved in the utilization of compounds classified as peptides (Fig. 5.1). Some of them included important microbial biomarkers, such as DL- $\alpha,\epsilon$ -diaminopimelic acid (DAP, Table C.6), occurring specifically in bacterial cell walls (Stevenson, 1994), and its analogs (e.g. 4-methylene diaminopimelate and (2*S*,3*R*,6*S*)-2,6-diamino-3-hydroxydiaminopimelate, Auger et al., 1996) found in both CW and HW non-resin treated extracts, as well as after the resin treatment of CW extracts. While their utilization was most likely involved in the reconstruction of cell wall peptidoglycans, L-lysine formation ensuing decarboxylation or a potential microbial inhibition by DAP analogues (Auger et al., 1996) can be considered in the WEOC<sub>min</sub> patterns observed. In spite of this, their presence confirms the microbial signature of the post-resin treated CW sub-fraction. However, it can only be speculated upon that such a CW sub-fraction derives from the microbial re-processing of dead cell components and debris, in contrast to the non-resin treated HW extracts, where these compounds would have been “artificially” released with high extraction temperatures. In fact, dialanine, a component of the peptidoglycan backbone (Hill et al., 2012), was detected, and utilized, in non-resin treated HW extracts only.

A large proportion (~56%) of compounds included in the peptide class and involved in biodegradation processes remained, however, unidentified, and only some estimated structural features were further analyzed.

#### 5.4.5 A closer look at the character of “not matched” compounds

Although the present study did not aim to quantify the contribution of different biochemical classes to the WEON pool, it is clear that a large proportion of WEON comprised the, hereby defined, not matched class (Fig. C.18 and C.19). Both non- and post-resin treated CW and HW extracts comprised of nucleobases (i.e. adenine, guanosine, cytosine, hypoxanthine, Table C.6). It was previously hypothesized that microorganisms release these compounds (*Bacillus sp.*) as a strategy for cross-

feeding of other heterotrophs to reduce substrate competition (Baran et al., 2015). This hypothesis is well aligned with the occurrence of e.g. adenine in both pre- and post-incubation non-resin treated HW extracts and its disappearance from post-resin treated HW extracts after the incubation. Baran et al. (2013) reported the ability of some bacteria (*E. coli* and *S. oneidensis*) to grow on adenine (as well as the other nucleobases here listed) as either a C or N source. Accordingly, lacking other energetically viable nutrient resources (e.g. amino-compounds and peptides), such as in post-resin treated HW extracts, microorganisms would have utilized nucleobases for their sustainment, with consequent release of CO<sub>2</sub> and N assimilation.

In contrast, some of these compounds were still found in the post-resin treated CW extracts. In support of this hypothesis is the presence of (*S*)-ureidoglycine (Table C.6), revealed only post-incubation, yet in both CW and HW post-resin treated extracts. (*S*)-ureidoglycine is a nonstandard  $\alpha$ -amino acid directly derived through the degradation of allantoate formed during purine catabolism (ureide pathway). Such a pathway aims at the mobilization of the ring N for further assimilation by plants (e.g. *A. thaliana*) and microorganisms (e.g. *E. coli*) (Serventi et al., 2010; Shin et al., 2012). The process releases CO<sub>2</sub>, as well as ammonia, as a by-product of allantoate degradation. While (*S*)-ureidoglycine alludes to the mineralization of nucleobases and N assimilation in the post-resin treated sub-fractions, it is questioned why would microbes then (partly) utilize nucleobases in the post-resin treated CW extracts comprising of readily degradable substrates, such as the amino-N pool? One possible explanation would assume microbial substrate specialization in order to avoid competition (Baran et al., 2015). In addition, McGill and Cole (1981) and Qualls and Haines (1992) suggested that N-containing compounds are not selectively utilized in response to N limitation, but C mineralization.

Other compounds, such as (sodiated) nitroacetamide (Table C.6), were mainly found in post-incubation samples of all extracts, suggesting that a proportion of not matched compounds might have, indeed, been replenished by degradation products. The ability of several fungi (Teramoto et al., 2004) and bacteria (Zhang and Bennett, 2005) to degrade nitro-aromatic compounds has been reported (Gilcrease and Murphy, 1995); thus it is likely that nitroacetamide derived from the reductive degradation of nitro-aromatic compounds and acetylation of the amine products and condensation.

Organic acids derived from the citric acid cycle (e.g. succinic acid and oxaloacetate) also comprised CW but also HW non-resin and post-resin treated extracts. Some, including *R*(-)-citramalic acid (Table C.6), were found after the incubation of only post-resin treated HW extracts from grassland and no-tillage soils. While some studies have indicated the release of citramalate from root exudates of pasture and crop legumes and grass species (e.g. Kidd et al., 2016), its post-incubation presence

in only post-resin treated HW extracts suggests its *de novo* formation, possibly from bacterial anaerobic metabolism of glutamate and cleavage of ammonia directed to the formation of ATP (Buckel and Barker, 1974) or the synthesis of isoleucine from an alternative pyruvate pathway by methanogenic archaea, involving acetate and CO<sub>2</sub> as C source (Ekiel et al., 1984; Howell et al., 1999). However, given that pyruvate was only detected in the non-resin treated HW extracts, where citramalate formation did not occur, its formation might be directly related to oxaloacetate (Table C.6) utilization (Tsai, 1967; Wang et al., 2019), a hypothesis that would need to be confirmed despite its disappearance observed after the incubations of the non- and post-resin treated HW extracts. Nevertheless, such a result alludes to the shortage of biodegradable C sources in the post-resin treated HW extracts and the potential utilization of CO<sub>2</sub>, which would help to explain the lower WEOC<sub>min</sub> observed in this sub-fraction.

It is evident that numerous metabolic pathways and sources, including heteroaromatic N compounds, define the nature and fate of soil WEOM and an integral commitment is needed to enable a clearer classification of metabolites and their identification as potential biomarkers in SOM studies.

#### 5.4.6 Inconsistent aspects of structural determinants

The unsaturation degree of WEOM was investigated by complimentary approaches. Given the strong dependency of DBE on N and P heteroatoms (Koch and Dittmar, 2006), C=C unsaturations were quantified by normalizing the DBE to the number of C atoms in a molecule (DBE/C, Koch and Dittmar, 2006). Further, the contribution of carbonyl unsaturations was assessed subtracting the O atoms from the molecule's DBE (Koch et al., 2005). Values of DBE/C were strongly biased by the relatively large N content. A threshold of  $\text{DBE/C} \geq 0.67$  was set for aromatic and highly condensed structures, yet leading to an overestimation of aromaticity due to heteroatom unsaturations (Koch and Dittmar, 2006). In such a context, the AI was developed to designate aromaticity through conservative constraints and a minimum value of 0.5, including all heteroatom contributions (Koch and Dittmar, 2006). However, the assignment of this parameter to infer aromaticity might lead to erratic conclusions. An example is given by adenine (Table C.6), detected in both non- and post-resin treated CW and HW extracts from all land uses, whose AI value would be undefined, or technically zero (Koch and Dittmar, 2006). In this case, the assumptions made by DBE/C would lead to a more appropriate interpretation, i.e. aromaticity, despite the relatively high values ( $\text{DBE/C} = 1.2$ ) driven by N in the cyclic array of  $\pi$ -bonds. The same conclusion is inferred by DBE-O, which ends up being a better indicator. Under these circumstances, adenine has been regarded as a highly unsaturated compound ( $\text{H/C} = 1$ ,  $\text{O/C} = 0$ ), yet not included in the classes defined by Rivas-Ubach et al. (2018), but its aromaticity

impact within a given system might have been overlooked.

In the present study, the investigation of the group of compounds not matching the classes defined according to Rivas-Ubach et al. (2018) involved the utilization of a modified AI ( $AI_{mod}$ ), expecting a large proportion of WEOM to be comprised of carboxyl groups (DiDonato et al., 2016). While  $AI_{mod}$  works well for compounds such as pyromellitic acid ( $AI_{mod} = 0.67$ , see Table C.6), whose aromaticity would have not been inferred by the conservative AI ( $AI = 0$ ), neither by DBE-O (DBE-O = 0), it is largely biased by N-C  $\pi$ -bonds (e.g. adenine) where, likewise AI, it fails to determine aromaticity. Furthermore, large overestimation may also occur. An example is given by (*S*)-ureidoglycine (Table C.6), erroneously classified as condensed aromatic molecule ( $AI_{mod} = 1.67$ ) in post-incubation, post-resin treated CW and HW extracts, or oxaloacetate (Table C.6) recovered in both non- and post-resin treated CW and HW extracts ( $AI_{mod} = 0.67$ ). In both cases, such overestimation was also reflected by DBE/C, while correct attributions were assigned by AI and DBE-O ( $AI = 0$  and DBE-O = -1 and -1.5, for (*S*)-ureidoglycine and oxaloacetate, respectively).

These inconsistencies led to a question as to what parameters might best represent the chemical and structural characteristics of WEOM that can explain  $WEOC_{min}$ . In view of the examples discussed, the general patterns depicted by the van Krevelen diagrams, and the range of values for the vast majority of compounds, all indicators were finally included in a PCA analysis, but patterns of their weighted average, rather than absolute values, were finally investigated. However, further work is clearly critical for an unambiguous characterization of WEOM.

#### 5.4.7 Relationship with WEOM biodegradability

Principal component analysis confirmed large differences in the optical and structural characteristics of non- and post-resin treated CW and HW fractions. Generally, bulk extracts were characterized by a larger degree of oxygenation (CW) and unsaturation (HW) with respect to the sub-fractions recovered after the resin treatment, thus revealing a shift towards lower MW compounds with higher H/C and lower O/C ratios falling into the lipid region (Ohno et al., 2010, Fig. 5.6), particularly in the HW extracts. These differences in the molecular structure of the post-resin treated WEOM sub-fractions were picked up by PARAFAC components 2 and 4, suggesting their possible olefinic and/or lipid-like structure, such as that of isoprene and fatty acid derivatives, but also nonprotonated aromatic cores, associated with quaternary aliphatic carbons, in contrast to condensed aromatic and phenolic-like compounds linked to components 3 and 5 of non-resin treated extracts. Leenheer et al. (2003) reported terpenoids (isoprene derivatives) as being DOM precursors refractory to biodegradation. In addition, as support for some of these hypotheses, an example of an interesting compound found in the post-

resin treated HW sub-fraction of the current study was 1,3-di-tert-butylbenzene, a volatile, refractory quaternary carbon, likely of microbial origin (e.g. rhizobacteria strain) (see Table C.6). To date, the fluorescence properties of such compounds and/or their interaction with surrounding fluorophores in DOM and WEOM studies have been largely overlooked by most general humic-like definitions and aromatic amino acids, and are, thus, understudied.

Despite their overlapping within the same region, post-resin treated CW extracts generally had higher oxygenation and lower unsaturation degrees compared to their HW counterparts, suggesting more labile lipid structures, such as those of short chain (<20) fatty acid derivatives (e.g. glaurin (Table C.6) or crotono-betaine, previously discussed), likely to be microbially derived (Lichtfouse et al., 1995; Pisani et al., 2014), but also a dilution effect from the larger proportion of labile N compounds. Albeit, the contribution of more stable quaternary carbon structures (e.g.  $C_{11}H_{21}NOS$ , Table C.6) or macrocyclic polyether compounds (crown ethers) capable of conjugation (e.g.  $C_{17}H_{27}NO_4S_3$ , Table C.6) was not excluded. With respect to the former, no evidence of crown ethers, their origin and involvement in DOM and WEOM studies has been found, but it is speculated these might have occurred as a result of humification processes (e.g. via synthesis of Schiff bases). There remained, however, in both post-resin treated sub-fractions, some exceptions of oxygen-free heteronuclei, such as nucleobases previously discussed.

There remain open questions as to what caused a comparable  $WEOC_{min}$  response in non- and post-resin treated HW extracts in lieu of the differences observed and what would have determined a larger  $WEOC_{min}$  in post-resin treated CW extracts given the relatively similar composition of CW extracts.

The variability in  $WEOC_{min}$  was strongly related to compositional differences between CW and HW extracts (Fig. 5.5), mainly determined by their post-resin treated sub-fraction. PARAFAC components were the most prominent factors of this separation, and in particular component 1, which was almost equally expressed in both non- and post-resin treated HW extracts in the order pasture>no-tillage>fallow (Fig. 5.5 and supporting Tables C.2 and C.3). This suggests that the largest fluorophore in the bulk HWEOM comprises of uncharged or protonated, unsaturated and poorly oxygenated moieties whose origin is strongly influenced by plants and associated microbial activity. These features, with the exception of the molecule's net charge, were in large contrast with those enhancing  $WEOC_{min}$ . It could be, therefore, concluded that an analogous effect resulted from the low oxygenation and high unsaturation degree of components 2 and 4, and the relatively more oxygenated, but largely aromatic components 3 and 5, likely related to the structures previously discussed. Although relatively low, the PARAFAC scores revealed by components 2 and 4 in the post-resin treated HW sub-fraction were, indeed, comparable or even higher than those revealed by components 3 and 5 in



the bulk HW extracts (Tables C.2 and C.3), suggesting that their impact on the  $WEOC_{min}$  is worthy of attention.

Interestingly,  $SUVA_{254}$ ,  $AI_{mod}$  and DBE/C had little if any effect on this separation, in contrast to DBE-O and  $abs_{254}$ , suggesting that the unsaturations inferred by the former were largely biased by the degree of oxygenation (i.e. C=O groups), which was much larger in the CW extracts. This is in line with the higher  $SUVA_{254}$  values observed in the non-resin treated CW extracts (see Chapter 4 and supporting Fig. C.1) and their progressive humification previously discussed. Another observation is the position of the C-peak, assigning a humic-like fingerprint to WEOC, likely derived from plant and SOM sources. This peak was consistently present in the non-resin treated CW and HW extracts, while the blue shift that was attributed to a transition (C-M) towards an M-peak (but not corresponding to an M-peak as defined by Coble, 1996) appeared mainly in the post-resin treated HW extracts, suggesting that: (i) there is an unexplained component (component 6?) underlying the C-peak, which is characteristic of CW extracts and masked by phenolic moieties; (ii) the shift towards shorter emission wavelengths (C-M) and higher H/C ratios designates the microbial origin of PARAFAC components 2 and 4 in post-resin treated HW extracts. However, while the position of the C-peak remained unaltered in the post-resin treated CW extracts, likely due to residual, poorly fluorescing structures, it can only be associated with enhanced  $WEOC_{min}$  for the low fluorescence intensities, not for its presence (Fig. C.2).

Yet, given the compositional similarity of these sub-fractions with the bulk CW extracts, it was suggested above that the resin-depleted fulvic fraction represented only a limited, but high C content CWEOC proportion. Assuming that this was the case, is it possible that such a smaller proportion of the CWEOM caused such large differences in  $WEOC_{min}$  (Fig. 4.1)? Even if the same labile C source (e.g. amino-compounds and peptides) was present in both fractions, and there was much less C per unit biomass in the post-resin treated sub-fraction, indicating that this source is only a small proportion of the total available C, it is clear that microorganisms likely elected different metabolic strategies to satisfy their needs. This would point at the utilization of only a preferred source according to elementary flux modes (Wortel et al., 2014), for which microorganisms will gain the highest energy production.

It was previously hypothesized (Chapter 4) that there was a close coupling of C and N in WEOM moieties of the sub-fraction remaining following the resin treatment and a regulatory role on the mineralization response as a function of secondary products from *de novo* synthesis or biodegradation processes. The proportional WEOC and WEON utilization (Fig. B.6) and the higher  $WEOC_{min}$  in the post-resin treated CW sub-fraction likely reveal the co-utilization of the sources remaining after the resin treatment. In contrast, the utilization of the non-resin treated CW fraction suggests that the most energetically viable source (e.g. peptides) is

utilized first, which would support the hypothesis (Chapter 4) that the  $WEOC_{min}$  of the CW extracts is mainly driven by its post-resin treated sub-fraction. Once an optimal microbial growth rate is established, they utilize the secondary sources, a phenomenon referred to as diauxie (Monod, 1947), which in this case appears to be effective in the pasture and no-tillage soils (Fig. 5.1), with consequent release of excess  $NH_4^+$  (Fig. B.4). In view of the recent findings of Wang et al. (2019), it may be that these sources (in the bulk extract) derive from a similar metabolic pathway, as opposed to the substrates isolated by the resin treatment, which may designate the utilization of sources from different metabolic pathways. It has already been shown that the latter case would lead to a higher microbial use efficiency than when utilizing those sources individually (Hermesen et al., 2015).

The same concept would then be applied to the patterns observed in the bulk HW extracts, confirming the hypothesis that small peptides (and other easily degradable sources, such as quaternary ammonium compounds) are prioritized within a large amount of available C. This is in line with the findings of Farrell et al. (2014), who demonstrated that the soil microbial uptake of peptides is driven by C rather than N excess. But these results also infer that phenols and other compounds falling under the not matched category, including organic acids and heteroaromatic N, are subsequently utilized (Fig. 5.1). The larger proportion of N with respect to the CW extracts led to a larger amount of excess  $NH_4^+$  released from microbial metabolism. There remains that the previously hypothesized, optically active and less oxygenated compounds of the post-resin treated HW sub-fraction are only partly and least utilized, mainly involving N-containing moieties, such as nucleobases or N-containing lipids (Gašparović et al., 2017). Indeed, the small fraction of amino-compounds classified under the peptide category in the post-resin treated HW extracts does not exclude a margin of error in the classification adopted (e.g. crotono-betaine), given the large overlapping of these compounds with lipid-like moieties on the van Krevelen diagrams.

## 5.5 Conclusions

Linking C and N cycles is crucial for predicting and controlling nutrient use efficiency, losses and environmental threats. This study builds upon the well recognized view of microbial metabolic adaption to the available substrate showing that the integral composition of the medium, rather than its fraction, may explain microbial utilization and the related mineralization response. Low molecular weight peptides and quaternary ammonium compounds were attributed an important role in regulating metabolic pathways of even more refractory, aromatic moieties, possibly serving as a readily available C source which then permits the mobilization of N from heterocyclic biomolecules. Although a large proportion of resolved molecules

remains undefined and an even larger proportion remains completely unknown, by means of fluorescence spectroscopy and structural determinants, such as heteroatom to C ratios, DBE and derived parameters, it was possible to estimate a substantial proportion of highly unsaturated and/or aromatic compounds participating in biodegradability processes. In spite of this, given the inconsistencies discussed around some of these aromaticity parameters, advancements in the prediction of this character are central to the interpretation of the ambiguous nature of WEOM, as well as the identification of new molecules, whose proportion in WEOM remains appreciably large. Here, the inclusion of Kendrick mass defects and complementary structural analyses, such as  $^{13}\text{C}$  solid state NMR and FTIR, in a framework for structural definition, would greatly improve the understanding of compositional changes.

PARAFAC of fluorescence EEMs enabled components of different nature to be distinguished, which would have been otherwise masked under the same peak region. For a large space of features, such as that of EEMs, estimating the model parameters at a given number of components greatly reduced the complexity of further data mining and interpretation. However, the current algorithm implementation requires more attention in regard to non-negativity constraints and processing tools for a robust data mining approach and a wider use. Nevertheless, the use of data mining has shown the ability to identify an appropriate number of components able to effectively describe WEOM's quantitative and qualitative characteristics commonly assessed by means of diverse analytical and spectroscopic methods (e.g. UV absorbance) and biochemical assays. It, thus, represents a promising and valuable tool capable of providing distinctive information on WEOM from different sources and fractions and ultimately, combining those with its biodegradability potential. The advantage of using data mining applications on rather powerful measures, such as fluorescence EEMs, is to reduce laborious and lengthy laboratory incubations and allow for rapid quantitative and qualitative characterization, while reducing the number of samples and analyses and ameliorating resource management. In this context, the inclusion of structural features might be an interesting tool for structural predictions and the characterization of broadly defined WEOM fluorophores.

The present work also showed that anion exchange was able to separate, with a certain level of confidence, a WEOM sub-fraction of microbial origin, likely to be comprised of low molecular weight zwitterionic compounds, quaternary carbon moieties, and relatively unsaturated, lipid-like structures. It is noteworthy that this method requires adjustments in order to account for pH differences among soil samples and its effect on molecule's reactivity. Further work should better characterize these compounds and investigate their involvement in biodegradability patterns and their interaction with the large suite of broadly defined, anonymous WEOM constituents. This could be achieved by performing isotope-labeling studies

to link the microbial communities involved to WEOM's metabolomic profile in order to gain valid understanding of formation and utilization pathways. An *in situ* study would also reveal the effect of sorption on soil minerals or aggregate preservation on microbial activity and community structure in response to variations in substrate availability.

## Supplementary material

Panel plots of the variables included in the data mining approach (Fig. C.1), EEM spectra averaged across each combination of land use, extracted fraction, resin treatment and pre- or post- incubation time (Fig. C.2–C.5), RF and LWL–GPD algorithm comparisons of RMSE (Fig. C.8–C.11) and model regressions (Fig. C.12–C.15), van Krevelen diagrams (Fig. C.16, C.17, C.18 and C.19), mean scores of PARAFAC components and weighted parameters included in PCA (Tables C.2, C.3, C.4 and C.5), PCA loadings (C.20), structural properties of some compounds (Table C.6), and LC–MS compound identification workflow are presented in Appendix C. In addition, the van Krevelen diagrams of reference extracts (i.e. peat and Suwannee River HA), which were not specifically discussed in the manuscript, are shown in Fig. C.21–C.25.

# Chapter 6

## Corollaries and Implications for future work

*“For the past century or two, nothing has mattered more for soils than their relations with human communities, because human action inadvertently ratcheted up rates of soil erosion and, both intentionally and unintentionally, rerouted nutrient flows.”*

John R. McNeill and Verena Winiwarter

Behind the scenes of soil drying and rewetting or land use effects, the present study clinched the role of DOM/WEOM composition on the ensuing biodegradability patterns. The suite of analytical methods adopted, ranging from more traditional oxidative and colorimetric methods, to more powerful spectroscopy and sophisticated chromatography and mass spectrometry, enabled the biodegradability profile of DOM and WEOM fractions, assessed by laboratory incubations, to be related to the characteristics of their components within a framework of prediction tools that do not rely on modeling assumptions. The objectives and outcomes have been described in the individual Chapters above; hereafter, a synthesis of the implications of these outcomes with respect to future work is discussed.

Chapter 3 showed that DOM properties vary between soil leachates and extracts when the soils have been kept at field capacity or air-dried. While air-drying strongly enhanced the bulk of the DOC released with respect to the other moisture treatments, advancing drought levels determined the rate of change in the amount of DOC sequentially leached: a steady decline in the grassland soil and inconsistent variations in the arable soil. These inconsistencies precluded the establishment of a relationship with DOC mineralization in the arable soil, whereas in the grassland soil, mineralization concurred with a progressively lower amount of DOC released with sequential soil leaching and an increase of the aromatic character, inferred

by means of specific UV absorption at 254 nm ( $\text{SUVA}_{254}$ ). Despite an overall fast short-term biodegradability of DOM, arable and grassland soils responded differently to the increasing proportion of aliphatic DOM released upon rewetting of progressively drier soil conditions. These results reiterate the role of soil structural changes, microbial interactions or changes in soil redox conditions brought about by different land uses in the resultant effect of soil moisture conditions on the dynamics of DOM release and mineralization. Further, they show the need for a better understanding of the biochemistry of DOM in relation to its reactivity and responsiveness to different land uses. Ultimately, they point to the need to invest more research in optimizing extraction methods and means of storing soil samples in order to avoid over- or under-estimations of DOM's size, minimize artifacts of its qualitative assessments and, therefore, abridge the still widespread discrepancies between laboratory and *in situ*, field experiments.

The results presented in Chapters 4 and 5 confirmed the usefulness of widely adopted fluorescence and UV spectroscopies in DOM characterization studies. Fluorescence fingerprint indices, most commonly utilized in studies on aquatic DOM, gained increasing interest for their application in soil DOM studies (e.g. Fernández-Romero et al., 2016), due to the correlation established between the exogenous DOM in such environments and DOM leached from soils, hence the general consistency of the values reported. In this study, values of FI corresponded to those originally found in the literature, including those from either river and wastewater isolates or compost and soil extracts (e.g. Chen et al., 2003; Antizar-Ladislao et al., 2006; Jaffé et al., 2008; Rodríguez et al., 2014; Fernández-Romero et al., 2016). This was likely due to the dependency of the FI on shifts in the peak emission wavelength, thus enabling a microbial signature to be assigned to the WEOM sub-fraction isolated by anion exchange resin treatment, as opposed to plant and highly decomposed SOM constituents characterizing the bulk of the WEOM. Supporting evidence for this distinction was further provided by LC-MS profiling, promoting fluorescence indices as being valid in soil DOM and WEOM research. Their value in providing a “snapshot” of the origin and the degree of WEOM humification obtained with different extraction methods, thus, enables comparisons between studies and likely overcomes methodological impacts on the fractions studied.

More controversial was the information gathered from  $\text{SUVA}_{254}$ , due to the variety of structural configurations involved. Nevertheless, the results presented in Chapter 3 clearly show large differences brought about by soil drying or sequential leaching, substantiating its power to summarize the abundance of resonant structures and conjugated systems. In Chapter 4, this was further confirmed by the strong positive correlation between  $\text{SUVA}_{254}$  and the shift to low intensity fluorescence emission at 465 nm excitation (Milori et al., 2002), characteristic of molecular entities with extended  $\pi$ -systems, but also heterosubstituted aromatic structures

forming H-bonds with the medium. These changes suggested a greater degree of decomposition of the CWEOM fraction with respect to the HWEOM fraction, the latter displaying a much larger content of phenolic C, but lower emission intensities at 465 nm excitation.

However, inconsistent relationships between  $\text{SUVA}_{254}$  and DOM biodegradability among leachates, drying levels or land uses call for a better designation of this indicator. This will help to avoid misleading interpretations. For instance, the results presented in Chapter 5 suggested that the correlation between  $\text{SUVA}_{254}$ , DBE/C and  $\text{AI}_{\text{mod}}$ , reputed as aromaticity indicators, was largely biased by the degree of oxygenation, as indicated by the impact of oxygen on the degree of unsaturation (DBE-O), involving carboxyl and carbonyl functional groups often, but not always, associated with aromatic structures (e.g. Sleighter and Hatcher, 2008). Emerging research has been emphasizing the presence of relatively stable, O-functionalized CRAM and quaternary carbon structures, likely derived from humification processes of plant residues and microbial debris (Hertkorn et al., 2006; Mao et al., 2007; DiDonato et al., 2016; Gao et al., 2016a; Cao and Schmidt-Rohr, 2018).

This may explain the inconsistent (Chapter 3) or unexplained (Chapter 5) association between  $\text{SUVA}_{254}$  and DOM mineralization. Therefore, the structural variability held under the absorbance at 254 nm needs to be complemented with more recent  $^{13}\text{C}$  NMR and MS findings, to fully leverage the information provided by this simple indicator in DOM and WEOM biodegradability studies. From this perspective, the more conservative  $a_{254}$  was consistently inversely related to patterns of mineralization (e.g. Chapter 5). Moreover, changes in  $a_{254}$  were strongly described by PARAFAC components (Chapter 5), suggesting that the primary fluorophores responsible for the changes in maximum fluorescence intensities within distinct regions of EEMs (viz. broadly defined C- and M-peaks, Coble, 1996) correspond to the components absorbing at 254 nm; an aspect further encouraging the utilization of fluorescence analyses as being more informative of compositional changes occurring during biodegradation or from environmental and management factors.

In Chapter 5, PARAFAC of EEMs was used as an input for testing several machine learning algorithms with the aim of identifying functionally meaningful components capable of describing the variability in physicochemical and biological properties of WEOM. A rather striking outcome was the power of the Random Forest algorithm to describe the interrelated patterns of WEOM quantity, quality and biodegradability, inferred from a number of different analyses and demanding laboratory incubations, using the PARAFAC components from the spectral decomposition of fluorescence EEMs. The algorithm inferred 70% of the variation in WEOC mineralization was described by six PARAFAC components, which were inversely associated with patterns of biodegradability and strongly reflective of a

high degree of unsaturation, characterizing especially the HWEOM. While its utilization in natural sciences for conducting biodegradability studies has only started to gain attention, its application as a means for predictions of DOM and WEOM biodegradability is novel. The results presented in Chapters 4 and 5, hence, show a rather promising tool for this purpose.

In addition, this algorithm denoted the relatively weak ability of WEOM concentration to describe changes in mineralization, supporting the idea that C availability does not necessarily imply its biodegradability (introduced in Chapter 3 and corroborated in Chapter 4). In Chapter 4, for instance, the larger (3 to 4-fold) amount of HWEOC compared to the CWEOC corresponded to comparable C mineralized (32–34%), whereas small differences in the quantity of bioavailable C remaining after the resin treatment resulted in significantly different mineralization responses (about 20% higher in the post-resin treated CWEOC sub-fraction). In contrast, the algorithm showed relatively strong evidence of a relationship between the  $A_{465}$  humification index (Milorí et al., 2002) and the biodegradability patterns of WEOM extracted from different SOM fractions and isolated by anion exchange resin treatment (Chapter 4). The resin treatment distinguished an uncharged/protonated, aliphatic, more recent and likely microbially derived WEOM sub-fraction, comprising of low molecular weight N compounds, but also quaternary carbon moieties and relatively unsaturated, lipid-like structures, from the aromatic, relatively humified and/or phenolic, lignin and SOM-derived components comprising the bulk extracts. These results indicated a significant contribution of the aliphatic post-resin treated sub-fraction to the overall mineralization response of the CWEOC, but not of the HWEOC, likely driven by the mineralization of its WEON pool as a C source in presence of less energetically viable compounds, such as phenolics. These findings suggest that even in presence of large concentrations of bioavailable C, microorganisms selectively utilize only the preferred source to gain the highest energy production (e.g. Wortel et al., 2014).

Taken in isolation, the association of these differences with WEOM mineralization was, however, not apparent. This suggested that, despite the valuable information gleaned from fingerprinting and classification tools, generalizing upon compositional classes might not be sufficient to predict the microbial response, inferred by the resulting mineralization products, given the molecular specificity of biological systems and the diversity of communities apt to respond, or not, to a given substrate. In this context, the irregular and heterogeneous structural definition of humic substances certainly does not facilitate microbial decisions oriented towards efficient enzyme allocation and energy production (De Nobili et al., 2001). Even less can be learned from the large proportion of DOM compounds whose nature is still poorly defined. The ubiquitous presence of N in the structure of the molecules resolved in the present study suggested, however, that mechanisms involved in



SOM and DOM/WEOM formation are strongly related to mechanisms of N preservation. The large proportion of compounds not matching any of the designated classes here investigated alluded to the role of lignin in SOM and DOM/WEOM formation. Lignin derivatives and N heterocycles may, hence, provide a clue as to new structural definitions and pathways of N mobilization, but more effort must be invested to elucidate their contribution in shaping the “unknown” and its fate.

Quaternary ammonium compounds and small peptides represented the most abundant WEON source in both CW and HW extracted fractions. These findings were in line with the most recent studies arguing their contribution to the soluble N pool has been largely overlooked (Farrell et al., 2011; Warren, 2013b; Meadows and Wargo, 2015). While the ubiquity of these compounds has been demonstrated and their chemical characteristics formalized, little is still known about their role in biogeochemical cycles of C and N. Limitations have certainly derived from the fact that most investigations on DOM and WEOM require *a priori* selection of standard laboratory assays or analytical techniques, which builds upon preconceptions on the chemistry of soil DOM and WEOM. In this scenario, the development of powerful analytical techniques, such as LC–MS, helps pave the way for metabolomic studies, which are clearly one avenue of future research.

Evidence suggested the greater proportion of labile N compounds and the lower degree of unsaturation characterizing the post-resin treated CWEOM sub-fraction was likely to be regulated by differences in land use (Chapter 5). An increasing number of studies has demonstrated the importance of belowground C inputs to SOM formation and mineralization (Kong and Six, 2010; Kätterer et al., 2011; Mazzilli et al., 2015; Sokol and Bradford, 2019). The results presented in Chapter 4, and in particular evidence of the land use effect on the mineralization of the post-resin treated CW sub-fraction (Subsection 4.1.3), are strongly in line with these findings and advance the importance of plant inputs in shaping the size and quality of microbial outputs, expanded upon in Chapter 5. Furthermore, for both CW and HW extracts, and in particular for CW extracts, an overall reduction of compounds observed in the fallow soil relative to the pasture and no-tillage soils (Chapter 5) confirms the importance of plant inputs in WEOM quantity and composition. Thus, an additional distinction between belowground and aboveground OM sources may provide valuable understanding of land use and management effects (Mazzilli et al., 2015) on DOM/WEOM formation and mineralization in relation to the qualitative indicators promoted by the present study (e.g.  $A_{465}$ ).

The research presented in this thesis grounds on long-term field sites, yet validation of these emerging biodegradability proxies can only be achieved while including them in both new and ongoing long-term field experiments, over different time scales and along the soil profile. This would further allow apt model parametrization for the prediction of long-term biodegradability drivers and patterns. In this context,

the alternative approach to such predictions introduced by the present study, in particular while being at its early stage, invites to explore its potential and value for future research. Next, the introduction of an anion exchange resin treatment as a practical and convenient tool for the identification and characterization of functionally and structurally distinct DOM/WEOM fractions requires further method refinement, including a better understanding of hydrophobic interactions in WEOM by anion exchange resins and the validation of its potential on different soils from different environments.

While attempts to identify changes in DOM/WEOM fractions focused on our predictions of CO<sub>2</sub> release from a molecular perspective, the findings presented in Chapter 5 disclose a much more complex reality. For example, low mineralization rates may have negative implications for plant nutrition; yet, it was observed that aromatics can be, in fact, mineralized, demonstrating that it is not just a matter of perceived recalcitrance. These findings urge a sharp understanding of the microbial “multitasking” capabilities within a given context and the interactions occurring within the suite of microbial co-workers in a given scenario, in order to predict metabolic pathways of available substrates and their allocation. Linking microbial community structure and enzymatic activity to the soil metabolomic profile is a compelling challenge that must be undertaken in order to build authentic information at the microscale for further macroscale applications, such as the development of models of nutrient cycling across ecosystems. For such purpose, the metabolomic profile described in Chapter 5 may contribute to the design of *ad hoc* experiments for both qualitative and quantitative assessments of some target WEOM sources by conducting tracer studies and validating the fluxes by means of some of the most recent models developed (e.g. Wang et al., 2019). However, the introduction of metabolomics and molecular characteristics for the development of models for the prediction of C dynamics lies in the use of collaborative efforts and multidisciplinary expertise. The current study shows an example.

# Appendix A

**Table A.1:** Multivariate linear regression for  $\text{SUVA}_{254}$  as a function of DOC concentration (Fig. 3.5) in leachates from an arable and a grassland soil at different  $\theta_g$  before rewetting.

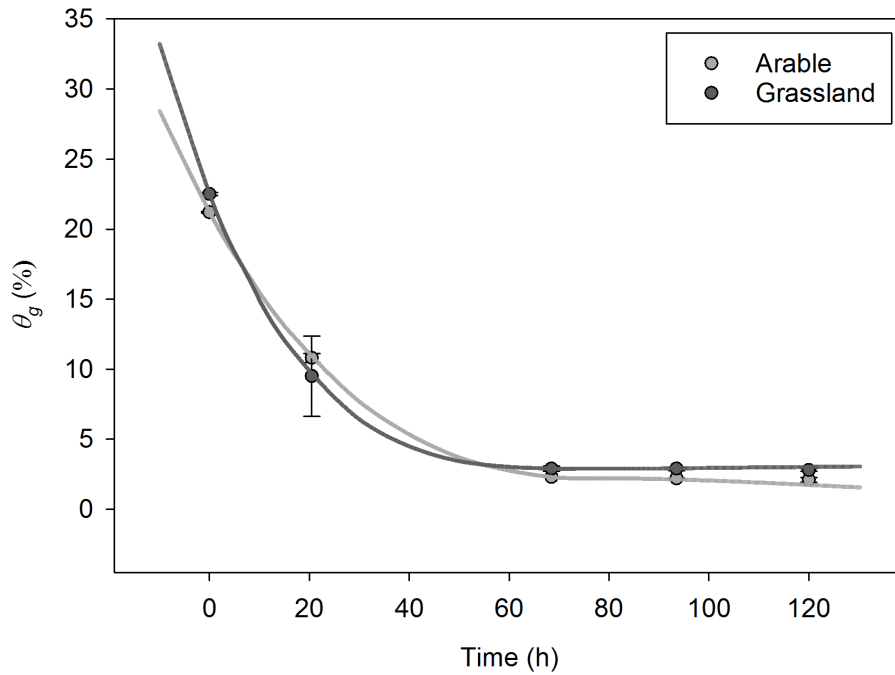
Land use	$\theta_g$ treatment	Intercept		Slope	
		Estimate	<i>P</i> -value	Estimate	<i>P</i> -value
Arable	FC	-1.86	0.388	15.96	<0.001
	15%	30.96	<0.001	-51.21	<0.001
	8%	9.21	<0.001	-8.88	<0.001
	AD	6.77	0.003	-3.29	<0.001
Grassland	FC	9.29	<0.001	-3.28	<0.001
	15%	9.07	0.677	-5.21	0.022
	8%	7.01	<0.001	-3.41	0.884
	AD	5.52	<0.001	-1.17	0.003

Model:  $y = y_0 + ax$ , where  $y_0$  = intercept and  $a$  = slope ( $R^2 = 0.63$  and  $0.85$  for arable and grassland, respectively)

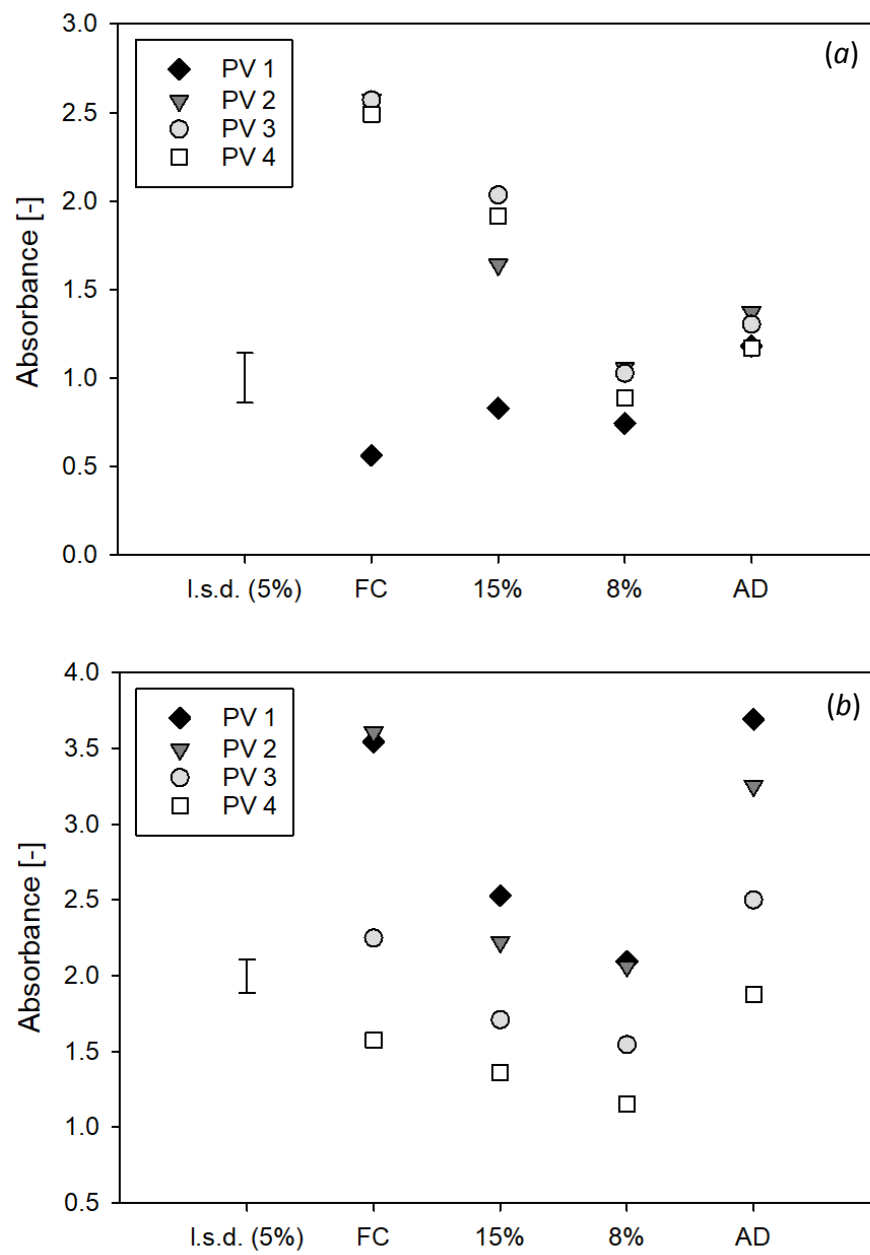
**Table A.2:** Multivariate linear regression for DOC biodegradability (DOC- $C_{min}$ ) as a function of DOC concentration (Fig. 3.8a) in leachates from a grassland soil at different  $\theta_g$  before rewetting.

Land use	$\theta_g$ treatment	Intercept		Slope	
		Estimate	<i>P</i> -value	Estimate	<i>P</i> -value
Grassland	FC	271.1	<0.001	-222	<0.001
	15%	186.1	<0.001	-124.4	0.001
	8%	209.9	0.001	-154.6	0.019
	AD	174.1	<0.001	-22.1	<0.001

Model:  $y = y_0 + ax$ , where  $y_0$  = intercept and  $a$  = slope ( $R^2 = 0.79$ )



**Figure A.1:** Experimental drying curves of an arable and a grassland soil utilized for the establishment of the designated  $\theta_g$  treatments at low temperature (4–6°C) and constant humidity. Equations:  $y = 2.01e^{-8}x^4 - 2.99e^{-5}x^3 + 7.35e^{-3}x^2 - 0.65x + 21.2$  ( $R^2 = 1$ ) and  $y = 3.14e^{-7}x^4 - 1.15e^{-4}x^3 + 1.54e^{-2}x^2 - 0.9x + 22.5$  ( $R^2 = 1$ ) for arable and grassland, respectively.



**Figure A.2:** UV absorbance at 254 nm of the pore volumes leached from an arable (a) and a grassland soil (b) at their designated  $\theta g$ . l.s.d. ( $\alpha = 0.05$ ) = 0.28 and 0.22. FC, field capacity; AD, air-dried; PV, pore volume.



**Figure A.3:** Arable site



**Figure A.4:** Grassland site





**Figure A.5:** Leachate collection from repacked arable and grassland soils.



**Figure A.6:** Infrared gas analyzer (IRGA) setup and gas sampling

This page is intentionally left blank



# Appendix B

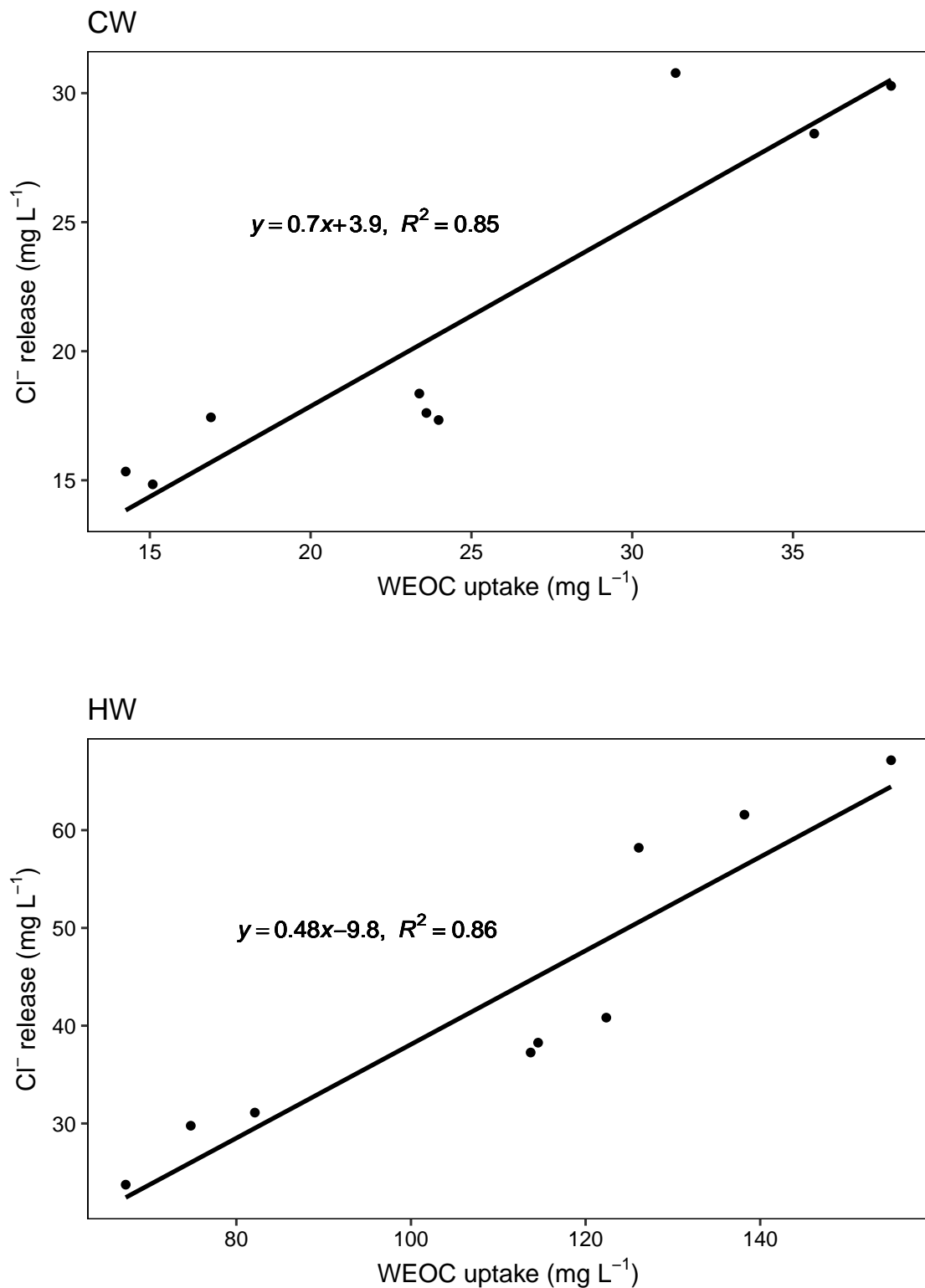
**Table B.1:** Concentration of anions in cold (CW) and hot water (HW) extracts before (a) and after (b) the resin treatment. Values are means ( $n = 3$ ); s.e.d = standard error of differences of means; l.s.d. = least significant difference ( $\alpha = 0.05$ ).

## (a) Non-resin treated extracts

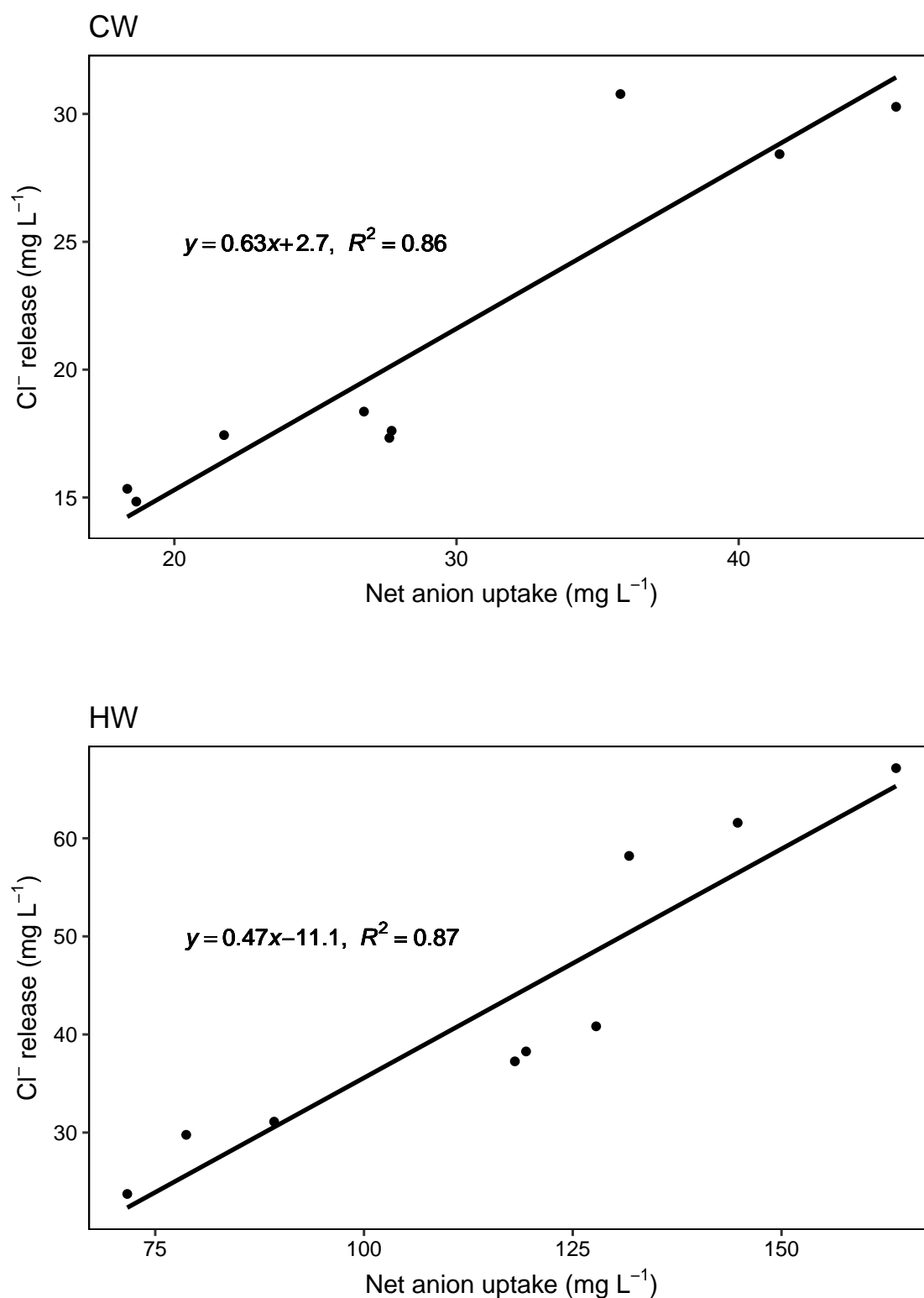
	Bicarbonate (meq L <sup>-1</sup> )		Nitrate (meq L <sup>-1</sup> )		Fluoride (meq L <sup>-1</sup> )		Sulphate (meq L <sup>-1</sup> )		Phosphate (meq L <sup>-1</sup> )	
Land use	CW	HW	CW	HW	CW	HW	CW	HW	CW	HW
Pasture	0.008	0.007	0.021	0.003	0.025	0.063	0.052	0.042	0.054	0.127
No-tillage	0.003	0.001	0.006	0.002	0.033	0.073	0.042	0.059	0.023	0.041
Fallow	0.003	0.001	0.014	0.002	0.013	0.040	0.048	0.045	0.024	0.079
s.e.d.	0.001	0.001	0.002	0	0.006	0.004	0.005	0.009	0.014	0.041
l.s.d.	0.002	0.002	0.004	0.001	0.016	0.012	0.013	0.026	0.040	0.114

## (b) Post-resin treated extracts

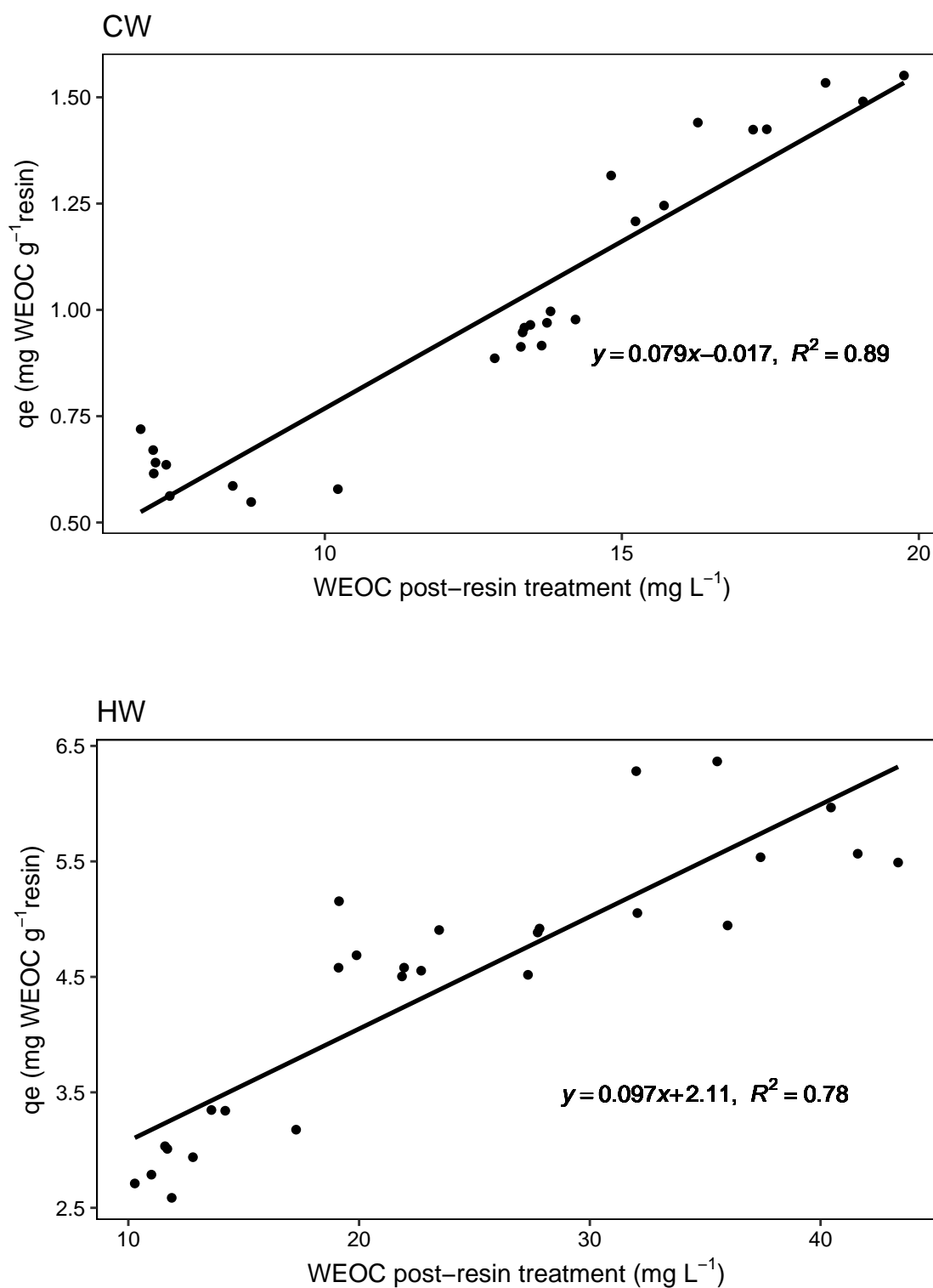
	Bicarbonate (meq L <sup>-1</sup> )		Nitrate (meq L <sup>-1</sup> )		Fluoride (meq L <sup>-1</sup> )		Sulphate (meq L <sup>-1</sup> )		Phosphate (meq L <sup>-1</sup> )	
Land use	CW	HW	CW	HW	CW	HW	CW	HW	CW	HW
Pasture	0.001	0.003	0	0	0.013	0.021	0.005	0.002	0	0.006
No-tillage	0.001	0.001	0	0	0.010	0.030	0.001	0.003	0	0
Fallow	0.001	0	0	0	0.006	0.018	0.001	0.002	0	0.024
s.e.d.	0.001	0	0	0	0.002	0.002	0.003	0.001	0	0.001
l.s.d.	0.002	0.001	0	0	0.005	0.005	0.009	0.003	0.001	0.003



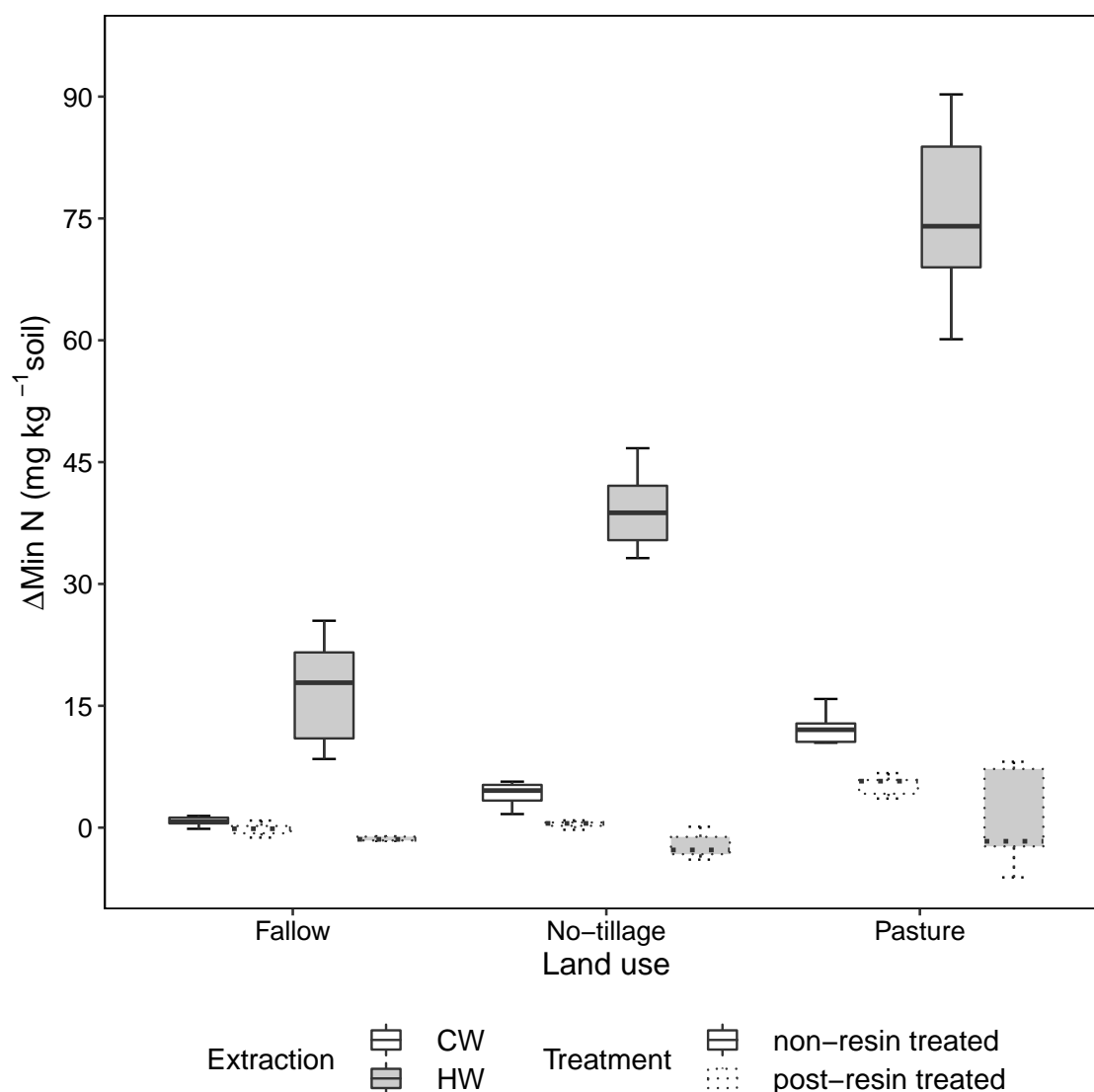
**Figure B.1:** Anion exchange stoichiometry of WEOC for chloride (Cl<sup>-</sup>) in cold (CW) and hot water (HW) extracts. Data points are means ( $n = 3$ ) of field replicates for each land use treatment.



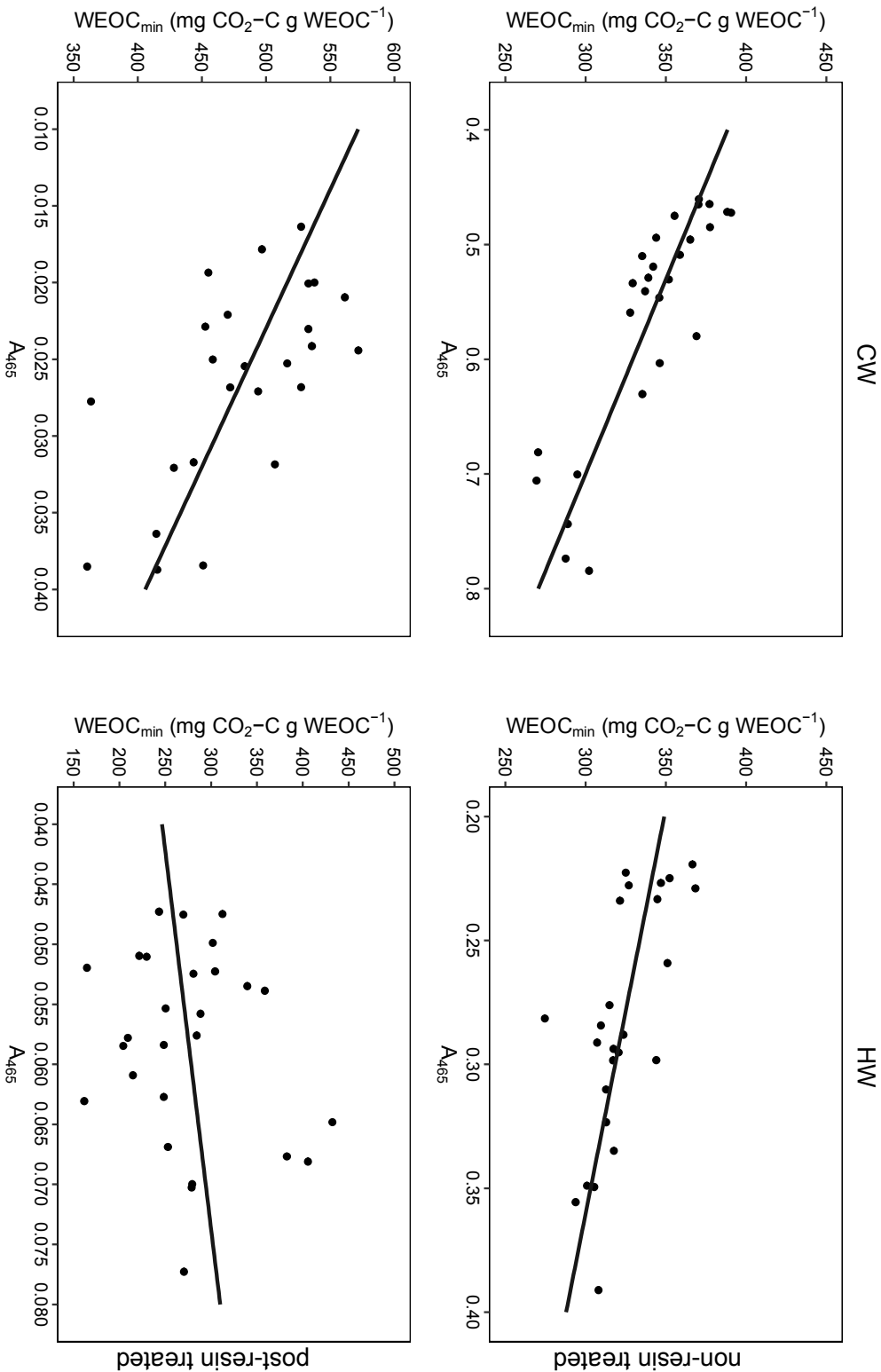
**Figure B.2:** Anion exchange stoichiometry of total anions (WEOC+HCO<sub>3</sub><sup>-</sup>-C+NO<sub>3</sub><sup>-</sup>+F<sup>-</sup>+SO<sub>4</sub><sup>2-</sup>+PO<sub>4</sub><sup>3-</sup>) for chloride (Cl<sup>-</sup>) in cold (CW) and hot water (HW) extracts. Data points are means ( $n = 3$ ) of field replicates for each land use treatment.



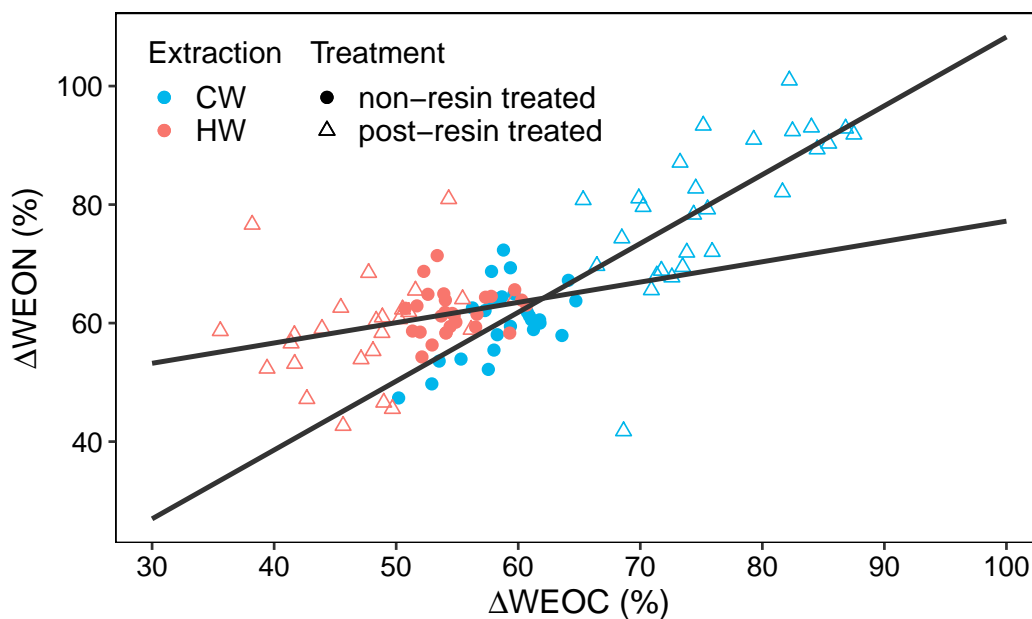
**Figure B.3:** Resin removal isotherms of cold (CW) and hot water (HW) extracts showing the equilibrium adsorption capacity ( $q_e$ ), estimated as the amount of adsorbed WEOC (mg) per unit mass resin (g), associated with the non-removable WEOC sub-fraction.



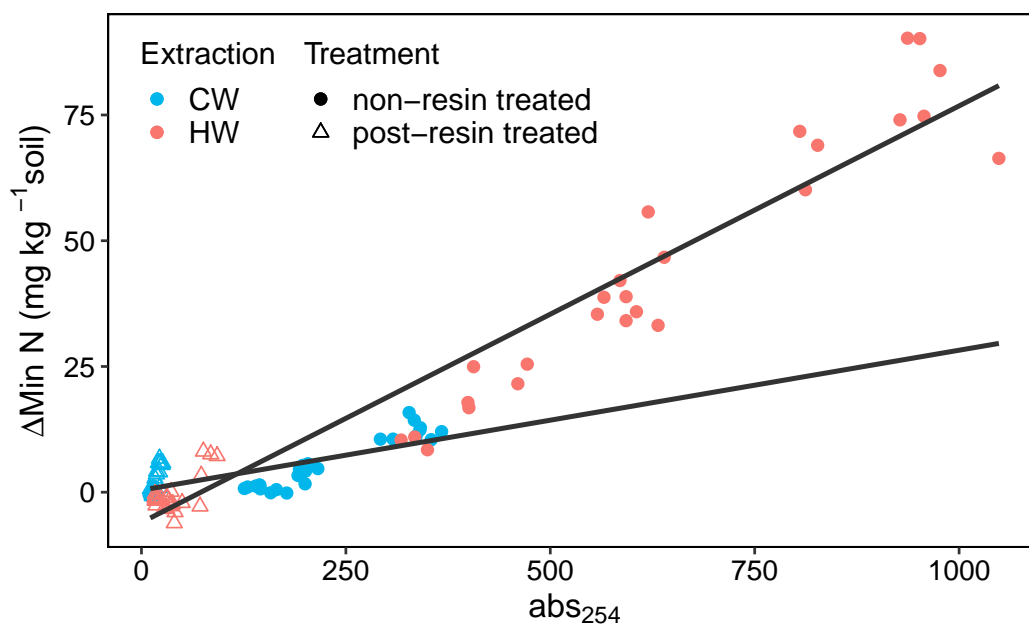
**Figure B.4:** Boxplots ( $n = 9$ ) showing the difference in mineral N ( $\text{Min N} = \text{NO}_3^- + \text{NH}_4^+$ ) before and after the 14-day incubation of non-resin treated and post-resin treated cold (CW) and hot water (HW) extracts of a soil under three different land uses. The upper and lower boundaries of the boxes show the 25<sup>th</sup> and 75<sup>th</sup> percentile, respectively; the central line represents the median; the whiskers indicate the minimum and maximum values.



**Figure B.5:** Random Forest regression analysis revealing the relationship between the A<sub>465</sub> Milori Humification Index and WEOC<sub>min</sub> in non- and post-resin treated cold (CW) and hot water (HW) extracts.



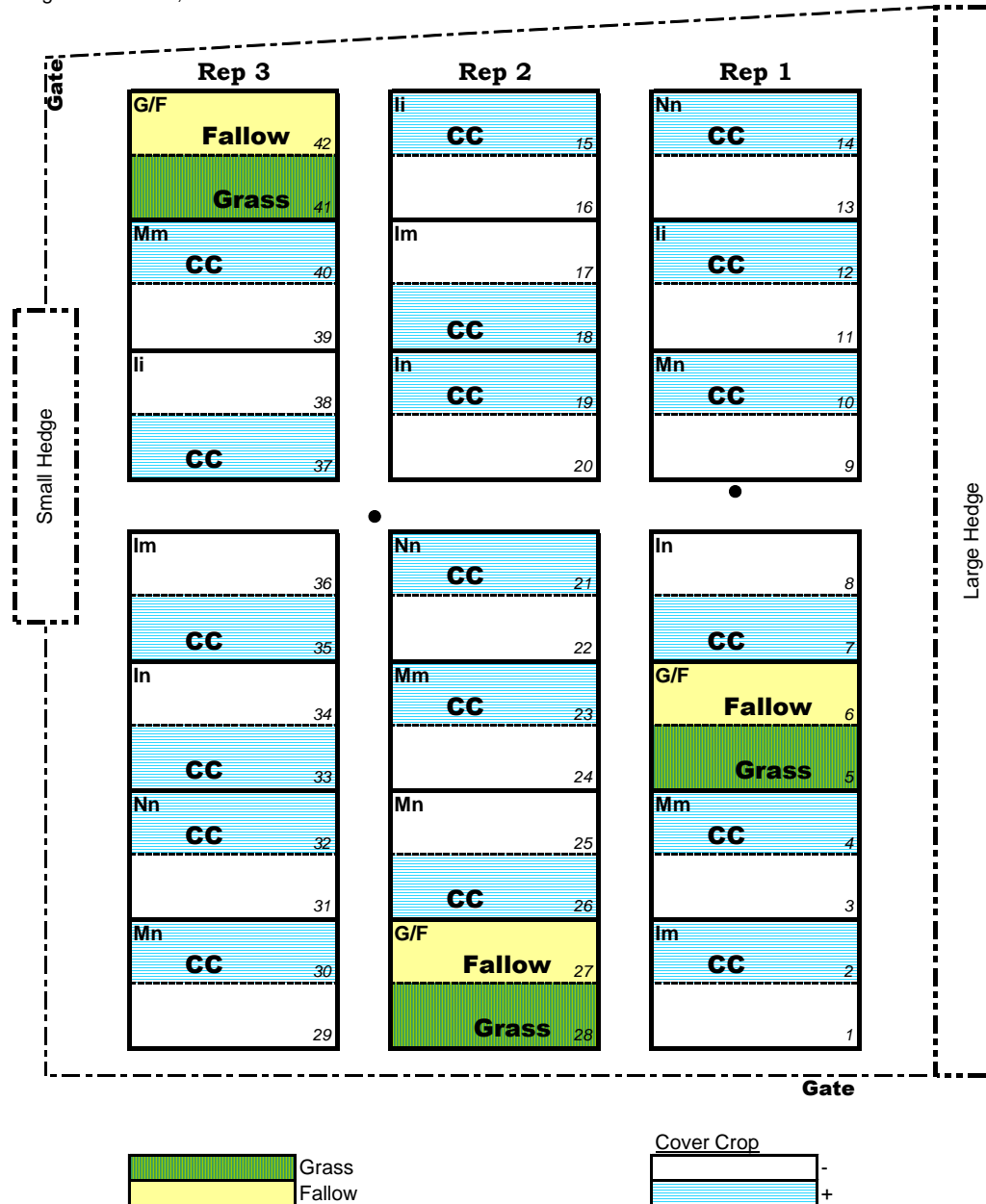
**Figure B.6:** Relationship between utilized WEOC ( $\Delta\text{WEOC}$ ) and WEON ( $\Delta\text{WEON}$ ) after the 14-day incubation of cold (CW) and hot (HW) extracts. Model:  $y = 1.16x - 7.89$  ( $R^2 = 0.71$ ) and  $y = 0.45x + 37.17$  ( $R^2 = 0.17$ ) for CW and HW, respectively.



**Figure B.7:** Relationship between the absorption coefficient at 254 nm ( $\text{abs}_{254}$ ) and mineral N ( $\text{Min N} = \text{NO}_3^- + \text{NH}_4^+$ ) for cold (CW) and hot (HW) extracts. Model:  $y = 0.03x + 0.41$  ( $R^2 = 0.57$ ) and  $y = 0.08x - 6.02$  ( $R^2 = 0.94$ ) for CW and HW, respectively.

### Millenium Tillage Trial

Design: Ruth Butler, Oct 2000



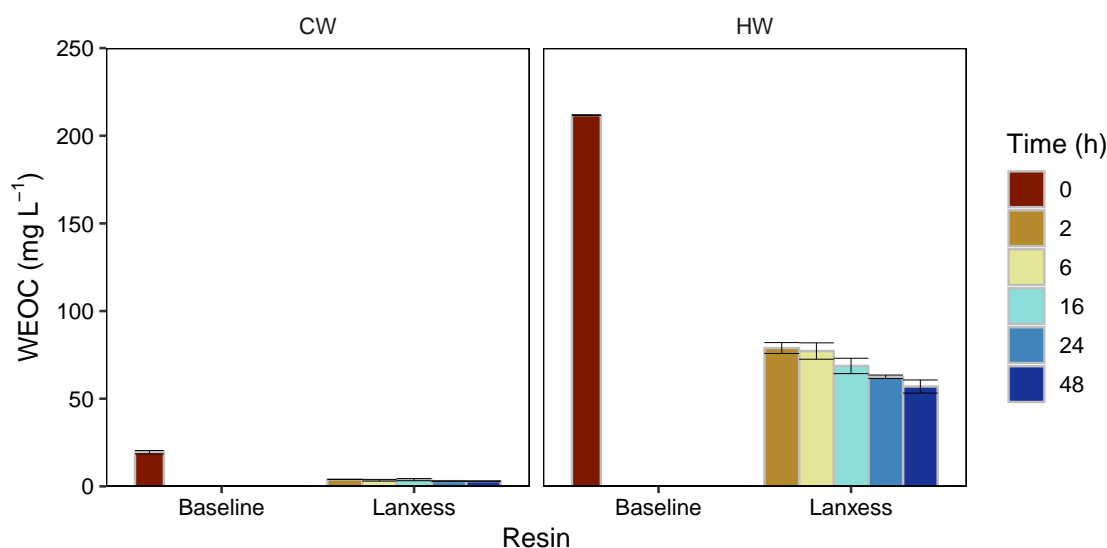
**Figure B.8:** Millenium Tillage Trial split-plot design with three replications for each land use treatment. The main plots are laid out in an incomplete Latin square (7 rows of 3 plots). Main plot treatment: cultivation; Split-plot treatment: cover crop. Cultivations (the first letter relates to spring time, the second letter to autumn): Nn: No-Tillage, No-Tillage; Mm: Minimum tillage, Minimum tillage; Ii: Intensive tillage, Intensive tillage; Mn: Minimum tillage, No-Tillage; Im: Intensive tillage, Minimum tillage; In: Intensive tillage, No-Tillage; G/F: Grass or Fallow.



## B.1 Anion exchange resin equilibration times

A batch equilibrium experiment was conducted on cold (CW) and hot water (HW) extracts from the MTT pasture soil to assess the anion exchange resin equilibration time allowing for the maximum WEOC removal. Triplicates of CW and HW extracts were added with the resin (Lanxess Ionac MacroT) at a resin:extract ratio of 1:20 and shaken on orbital shaker (80 rpm) for 2, 6, 16, 24, 48 h. The supernatant was then filtered through a 0.45- $\mu\text{m}$  membrane filter and assessed for TOC (Shimadzu Corp.).

The results demonstrated that, for the CW extracts, the equilibrium was achieved after 2 h, with  $\sim 80\%$  of WEOC removed (Fig. B.9). There was no significant WEOC uptake after 48 h ( $P = 0.11$ ), when the maximal removal was 85%. In contrast, WEOC removal significantly increased from  $\sim 63\%$  after 2 h up to 73% after 48 h in the HW extracts ( $P = 0.007$ ). These equilibration times were accounted for in the experiment described in Chapter 4.



**Figure B.9:** Determination of resin equilibration times and maximal WEOC removal from cold (CW) and hot water (HW) extracts. Baseline, initial WEOC concentration; Lanxess, WEOC concentration following Lanxess anion exchange resin treatment.

## Common Description

Delivery form	Cl <sup>-</sup>
Functional group	Quaternary ammonium; type 1
Matrix	Acrylic
Structure	Macroporous
Appearance	White, opaque

## Specified Data

Uniformity coefficient		max.	1.8
Range of size for >90 vol% of all beads		mm	0.40-1.60
Effective size	d10	mm	0.50-0.65
Total capacity (delivery form)		min. eq/L	0.85

## Operation

Operating temperature		max. °C	80 (Cl)
Operating pH range	during exhaustion		0-12
Bed depth for single column		min. mm	800
Back wash bed expansion per m/h (20°C)		%	10
Specific pressure loss kPa*h/m <sup>2</sup> (15°C)		kPa*h/m <sup>2</sup> (15°C)	1.1
Max. pressure loss during operation		kPa	250
Specific flow rate		max. BV/h	5
Freeboard	during backwash	min. vol. %	80-100

## Typical Physical and Chemical Properties

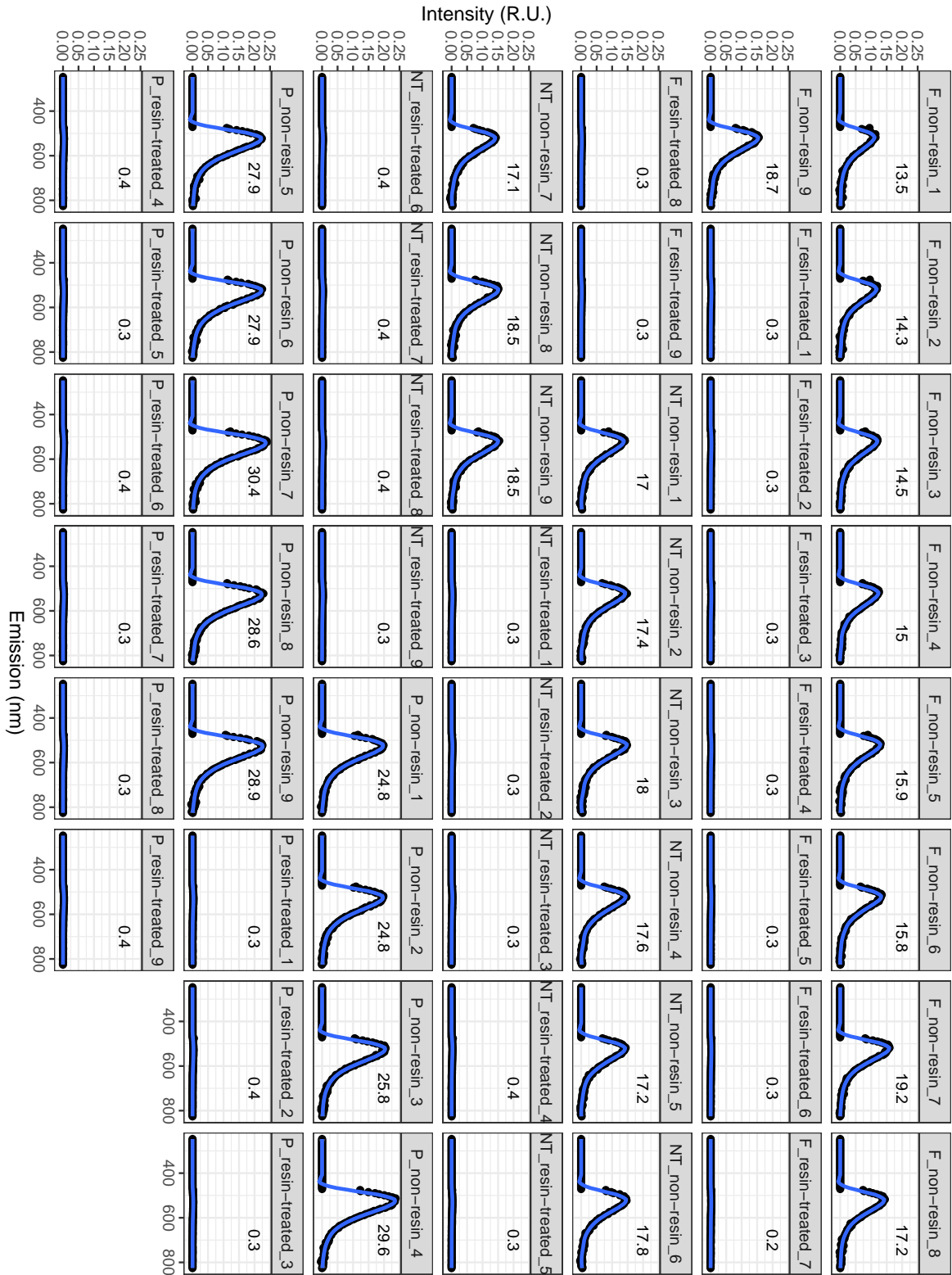
Bulk density for shipment	(+/- 5%)	g/L	720
Density		approx. g/mL	1.07
Water retention (delivery form)		approx. weight %	63-71
Volume change (Cl <sup>-</sup> -OH <sup>-</sup> )		max. approx. %	25
Stability pH range			0-14
Stability temperature range		°C	1-80 (Cl)
Storage time (after delivery)		min. years	2
Storage temperature range		°C	-20 - +40

**Figure B.10:** Properties of the anion exchange resin used in this study. Reprinted from Lanxess.

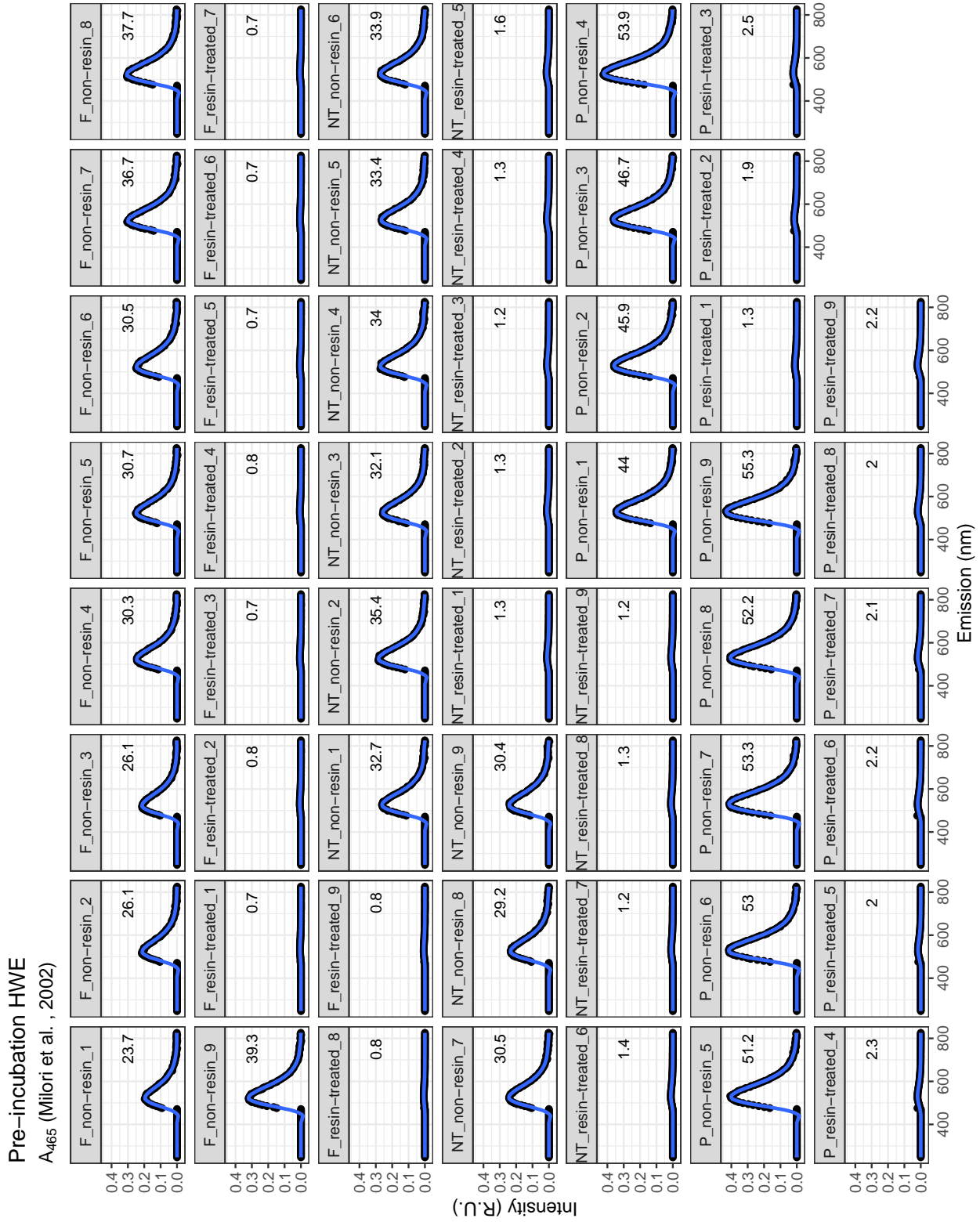
## B.2 Selection and calculation of the humification index

In order to select the appropriate humification index for this study, a preliminary inspection of fluorescence EEMs was handled within the R environment (R Core Team, 2018). The humification index (HIX) was calculated according to (Ohno, 2002) using the ‘eemR’ package (Massicotte, 2017). The  $A_{465}$  (Milor et al., 2002) was determined by fitting a locally-weighted regression (LOESS) smoother (Cleveland, 1979) to the emission intensities at an excitation of 465 nm, then integrating the fit across the entire emission range (247.2 to 825.5 nm). The smoothing span in LOESS was determined automatically for each sample via a repeated k-fold cross-validation (CV) procedure; here, 10 repetitions of 10-fold CV were elected (Kohavi, 1995) and run via the ‘caret’ package (Kuhn, 2008, 2015).

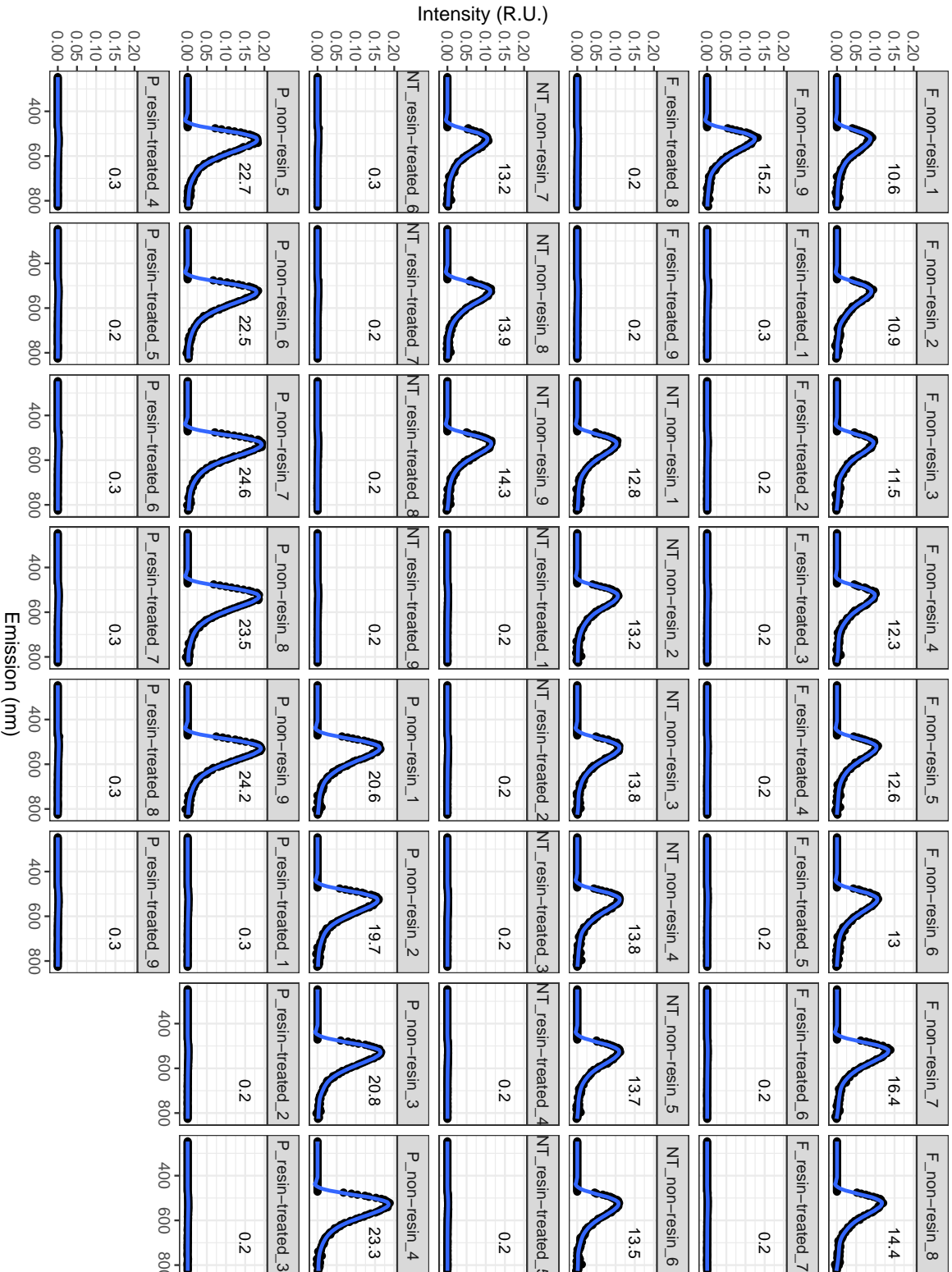
Values of HIX were disregarded (data not shown), given the lack of fluorescence emission at 255 nm excitation. Thus, only the results for  $A_{465}$  were summarized for pre- and post-resin treatment of CW and HW extracts (Subsection 4.1.2) and plotted as approximated smoother fits of the actual data via the `geom_smooth` function of the ‘ggplot2’ package (Hadley, 2016).



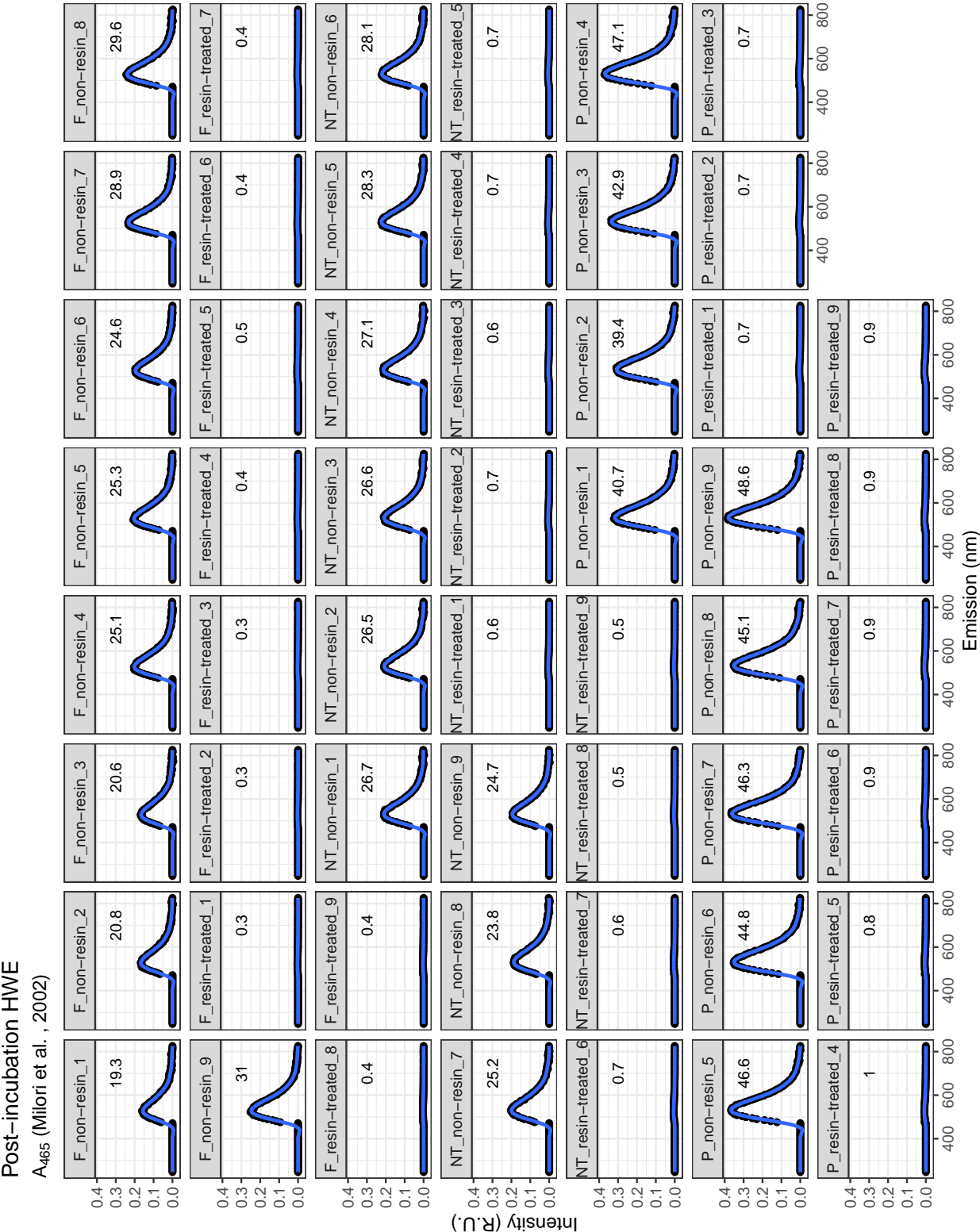
**Figure B.11:** Smoother fits and actual values for the A<sub>465</sub> Miori Humification Index in non- and post-resin treated cold water extracts (CWE). Intensities are in Raman units (R.U.). F: Fallow, NT: No-Tillage, P: Pasture, 1-9: sample replicates.



**Figure B.12:** Smoother fits and actual values for the A<sub>465</sub> Milori Humification Index in non- and post-resin treated hot water extracts (HWE). Intensities are in Raman units (R.U.). F: Fallow, NT: No-Tillage, P: Pasture, 1-9: sample replicates.



**Figure B.13:** Smoother fits and actual values for the A<sub>465</sub> Miliori Humification Index in non- and post-resin treated cold water extracts (CWE) after the 14-day incubation. Intensities are in Raman units (R.U.). F: Fallow, NT: No-Tillage, P: Pasture, 1-9: sample replicates.

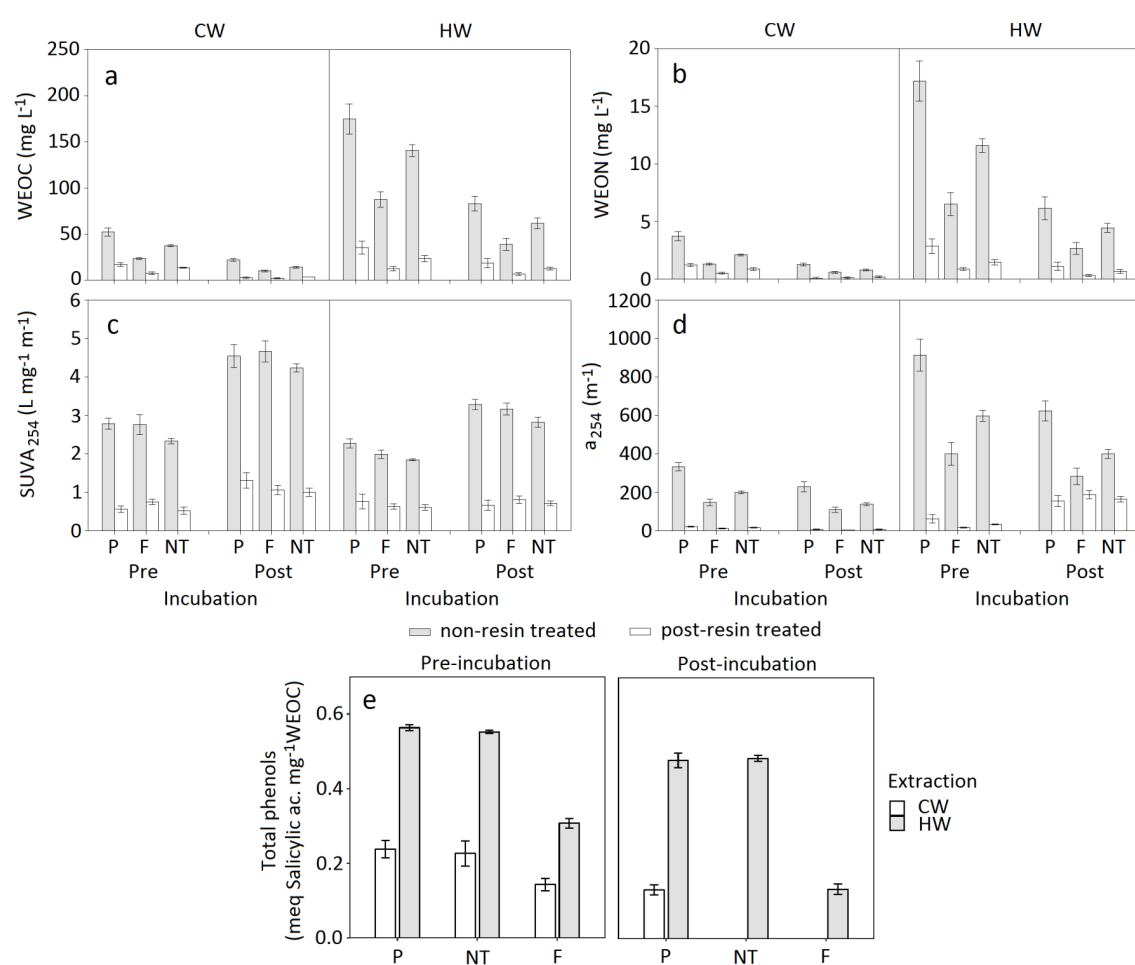


**Figure B.14:** Smoother fits and actual values for the A<sub>465</sub> Milori Humification Index in non- and post-resin treated hot water extracts (HWE) after the 14-day incubation. Intensities are in Raman units (R.U.). F: Fallow, NT: No-Tillage, P: Pasture, 1-9: sample replicates.

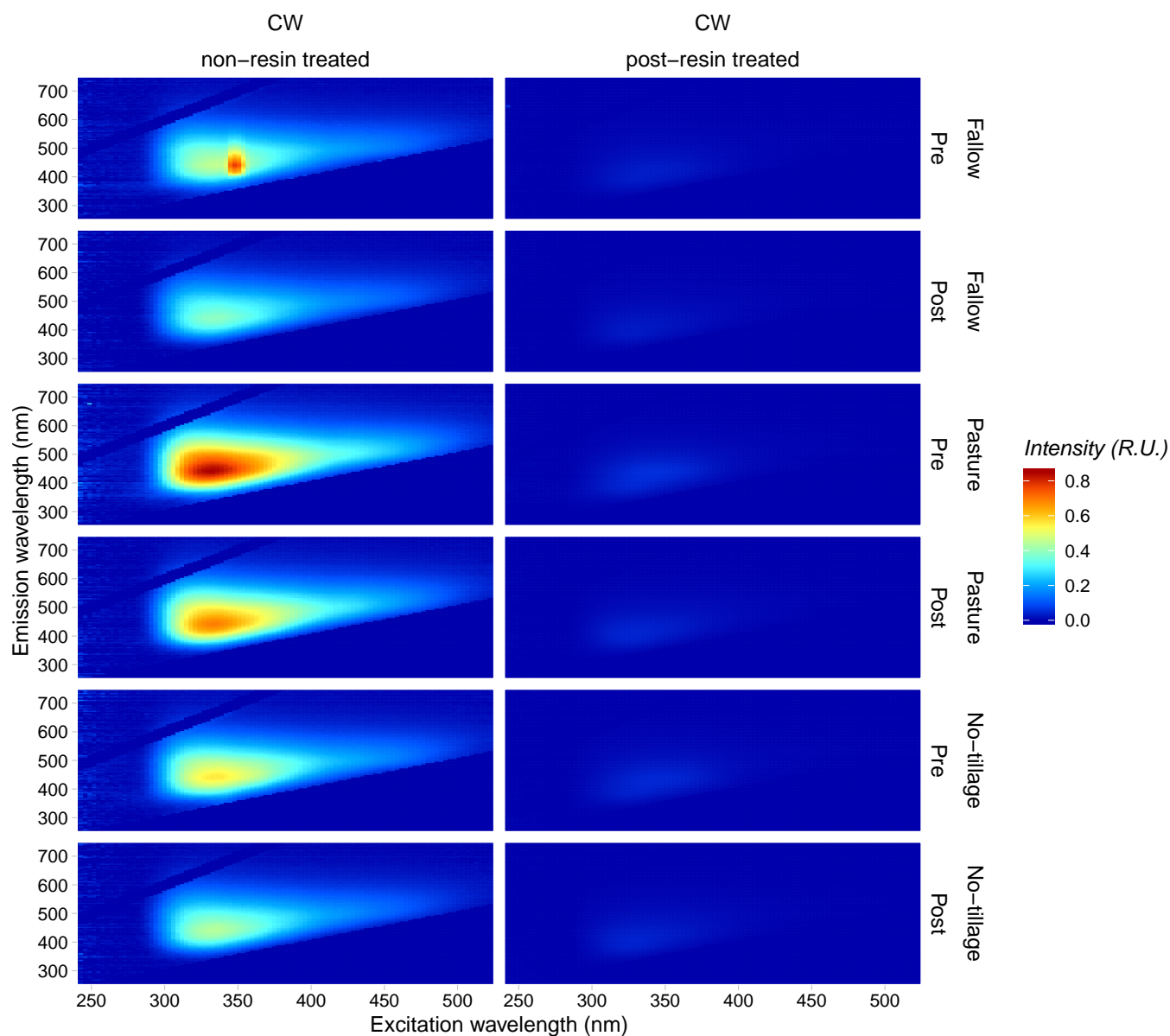
This page is intentionally left blank



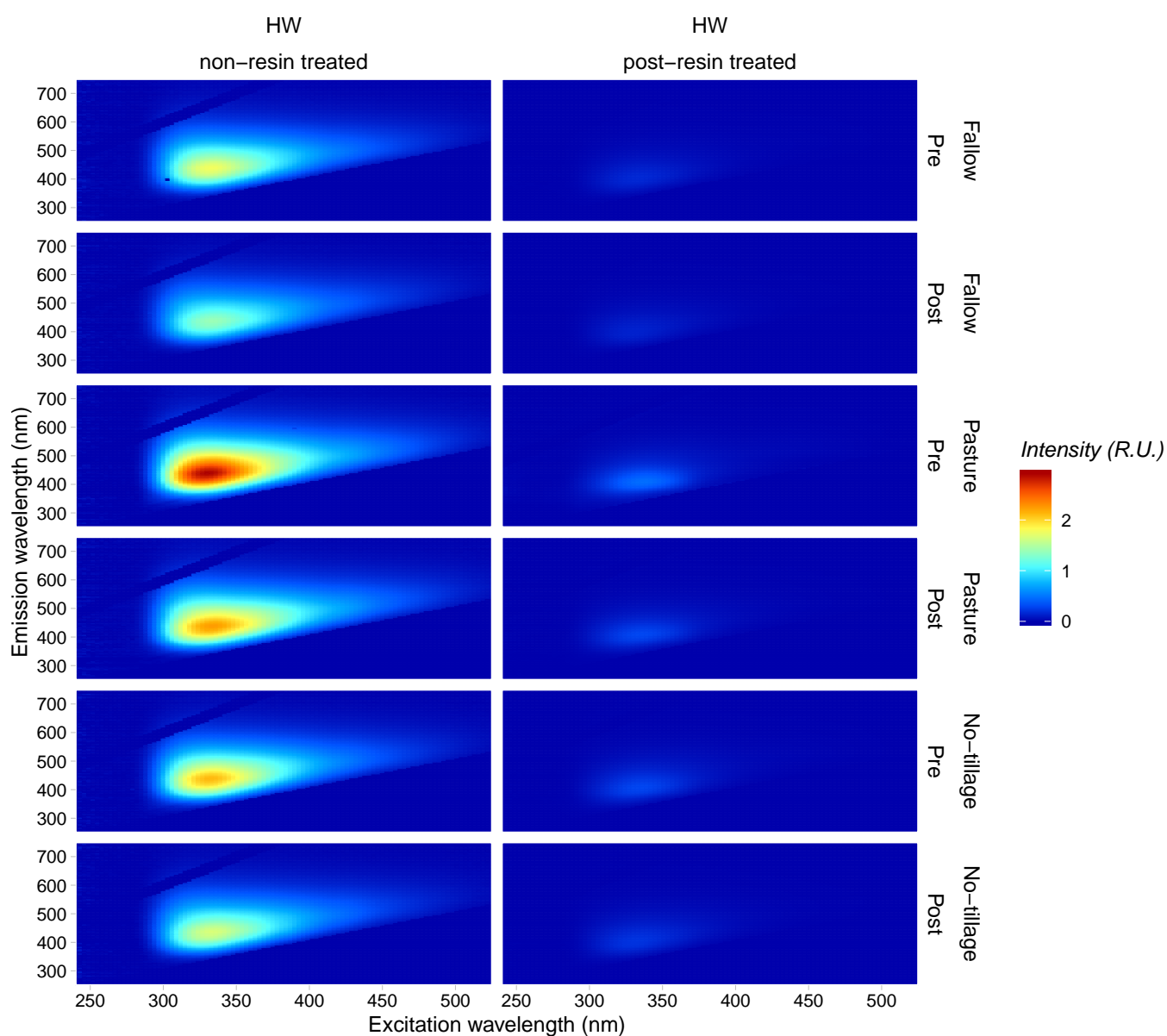
# Appendix C



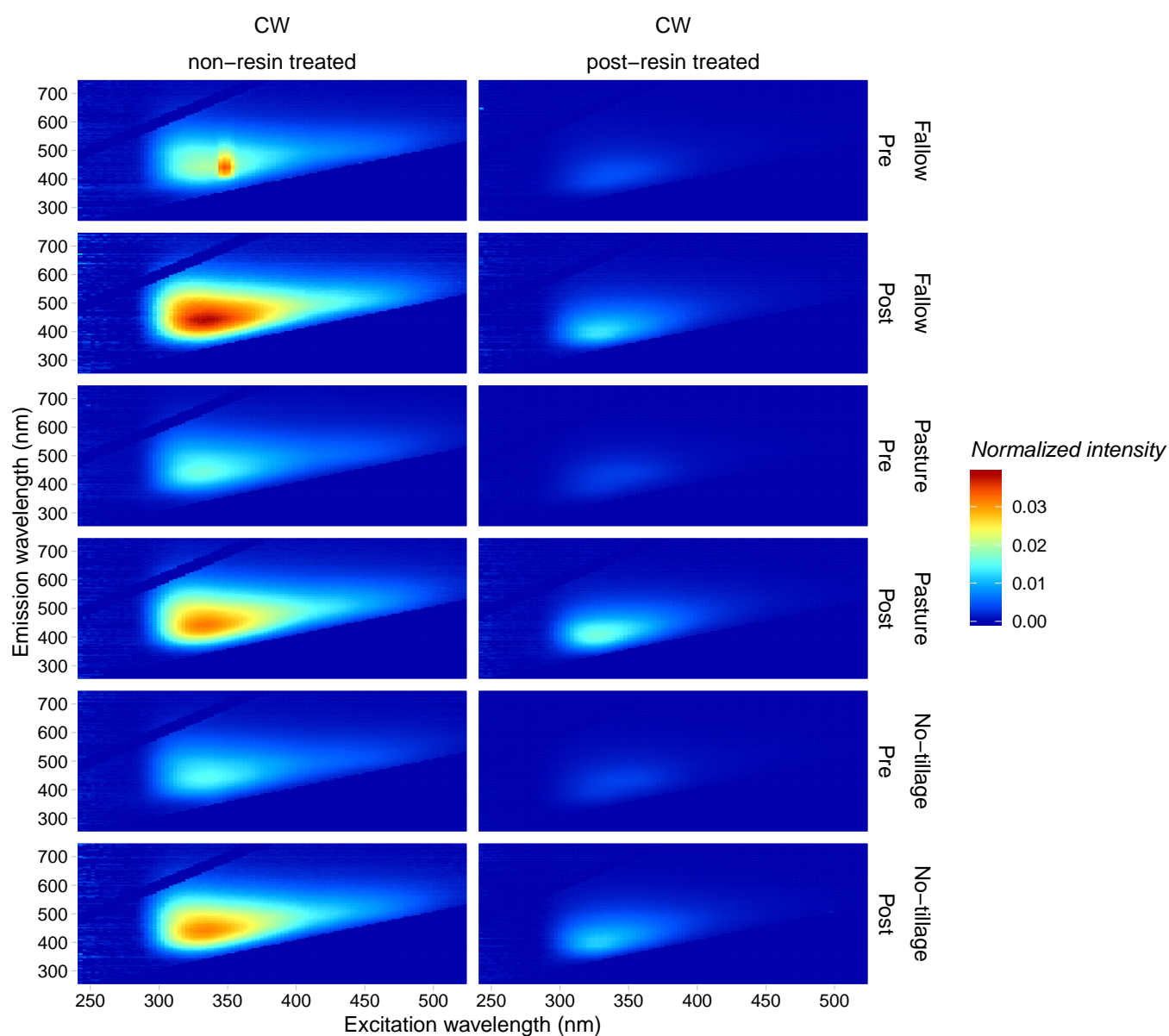
**Figure C.1:** Variables included in the data mining approach, showing pre- and post-incubation changes in non- and post-resin treated cold (CW) and hot water (HW) extracts of a soil under different land uses (Fig. a-d). Phenols (Fig. e) were not detected in post-resin treated samples. Bars are means ( $n = 9$ )  $\pm$  s.d.; P, pasture; F, fallow; NT, no-tillage.



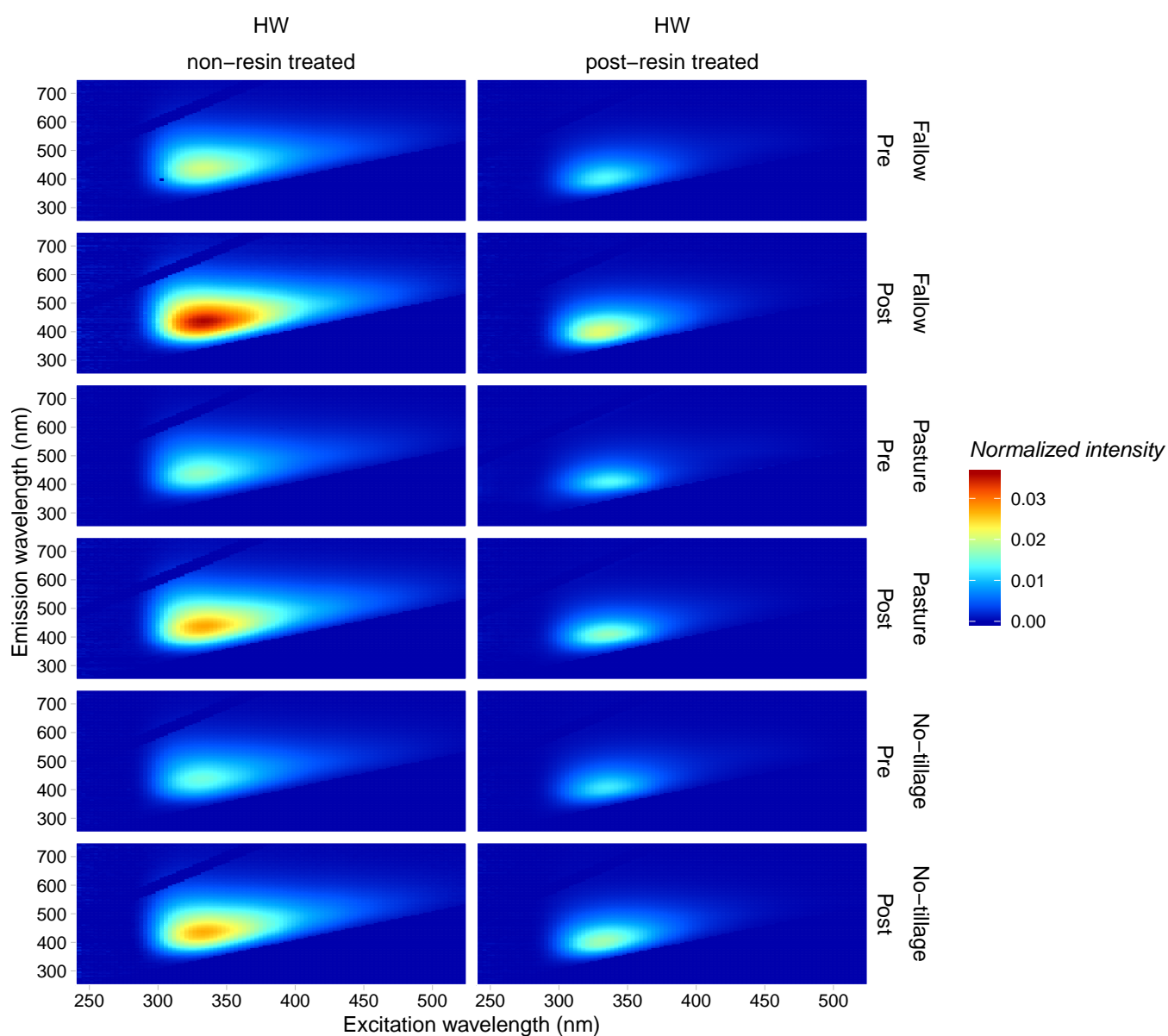
**Figure C.2:** EEMs of WEOM in non- and post-resin treated cold water (CW) extracts of a soil under three different land uses before (pre-) and after (post-) the 14-day incubation. Each EEM represents an average across nine replicates for each combination of land use, extracted fraction, resin treatment and incubation time. Intensities are expressed in Raman units (R.U.).



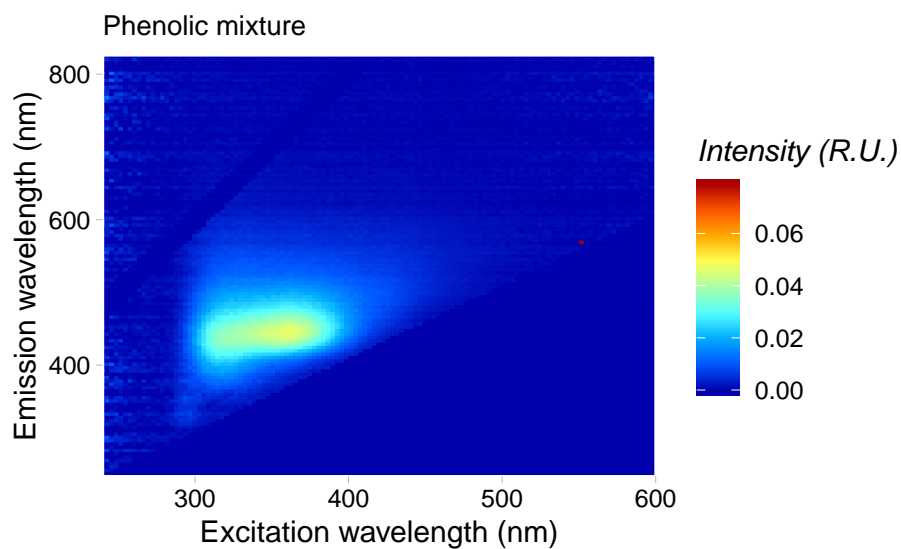
**Figure C.3:** EEMs of WEOM in non- and post-resin treated hot water (HW) extracts of a soil under three different land uses before (pre-) and after (post-) the 14-day incubation. Each EEM represents an average across nine replicates for each combination of land use, extracted fraction, resin treatment and incubation time. Intensities are expressed in Raman units (R.U.).



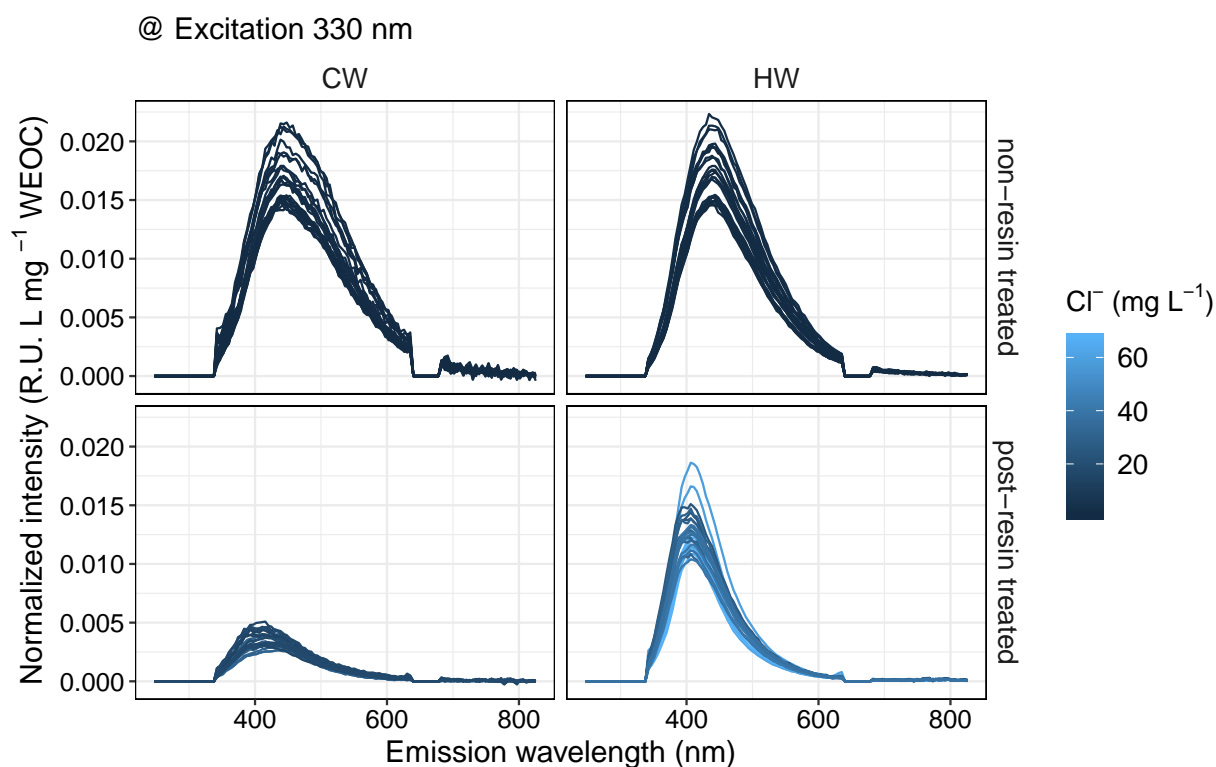
**Figure C.4:** Pre- and post-incubation changes in WEOM fluorescence of non- and post-resin treated cold water (CW) extracts of a soil under three different land uses. Each EEM represents an average across nine replicates for each combination of land use, extracted fraction, resin treatment and incubation time. Intensities are expressed in Raman units (R.U.) normalized to WEOC concentration ( $\text{mg L}^{-1}$ ).



**Figure C.5:** Pre- and post-incubation changes in WEOM fluorescence of non- and post-resin treated hot water (HW) extracts of a soil under three different land uses. Each EEM represents an average across nine replicates for each combination of land use, extracted fraction, resin treatment and incubation time. Intensities are expressed in Raman units (R.U.) normalized to WEOC concentration ( $\text{mg L}^{-1}$ ).



**Figure C.6:** EEM of the post-incubation phenolic mixture. Intensities are in Raman units (R.U.).



**Figure C.7:** Fluorescence emission at 330 nm excitation with varying chloride (Cl<sup>-</sup>) concentrations before and after the resin treatment.

## C.1 PARAFAC data mining

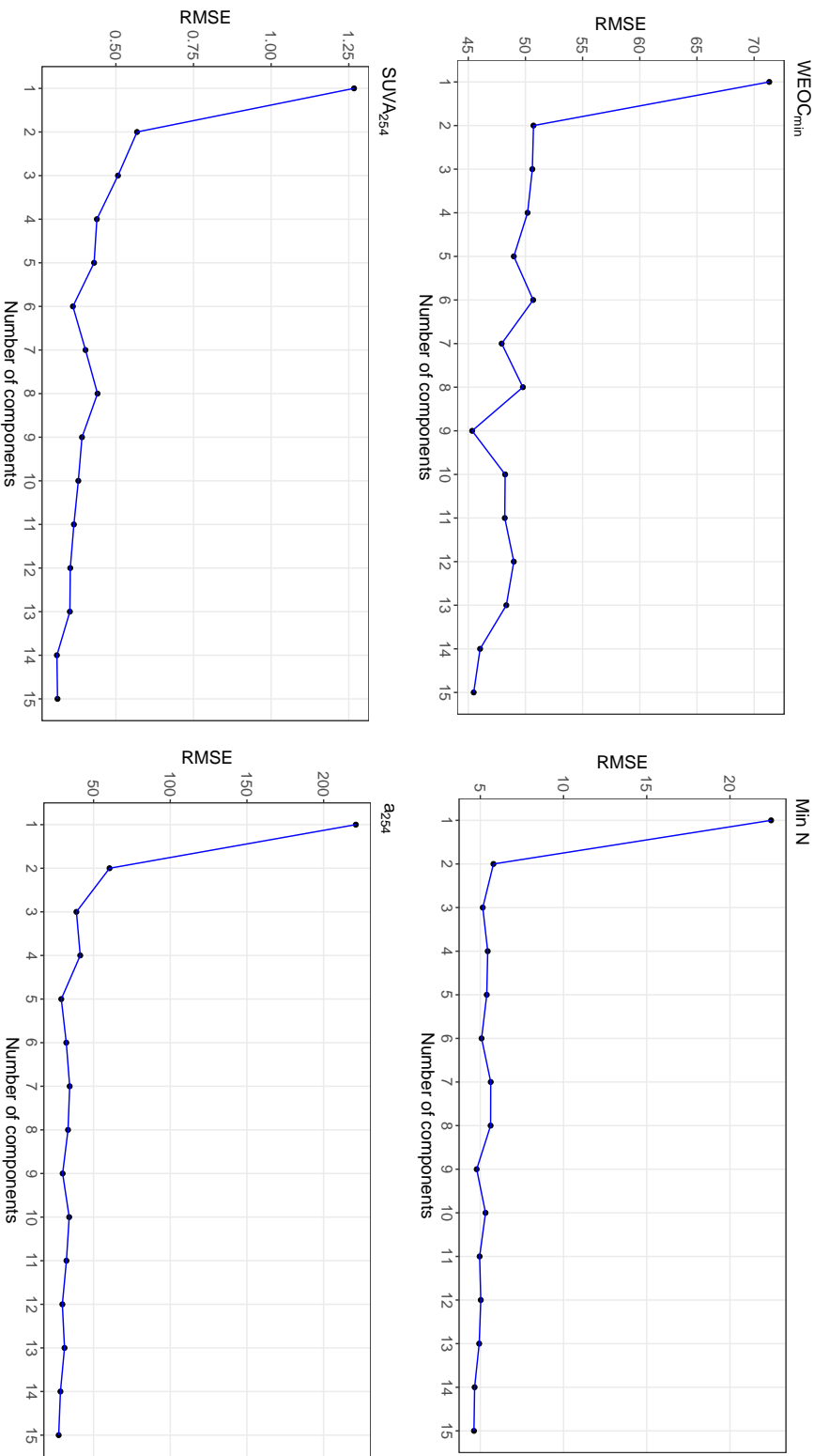
The selection of meaningful fluorescence components derived from the PARAFAC spectral decomposition included the evaluation of: RMSE, shape of the data points, Pearson correlation coefficient ( $r$ ) and coefficient of determination ( $R^2$ ) of the regression (described in Chapter 5) in relation to the variables of interest. Mainly two algorithms, Random Forest (RF) and Locally weighted Gaussian Processes (abbr. LWL-GPD), competed in the data mining of PARAFAC components.

For both algorithms, an initial screening of the RMSE revealed a substantial reduction explained by five or six components, followed by an increase (for most variables) preceding another large decrease at around eight or nine components (Fig. C.8–C.11). Given the wide variety of experimental factors (land use, extracted fraction, resin treatment, pre- or post-incubation analysis), a larger number of components, reflecting the variability in peak positions due to meaningful variability in the molecules present, shall not be excluded. Thus, a six vs. nine component model was evaluated (Fig. C.12–C.14).

For both models, overlaying values estimated from six or nine components showed minimal improvement of the model fit, suggesting overfitting when more than six components were used. Six PARAFAC components, hence, represented a meaningful set for the interpretation of WEOM properties and biodegradability.

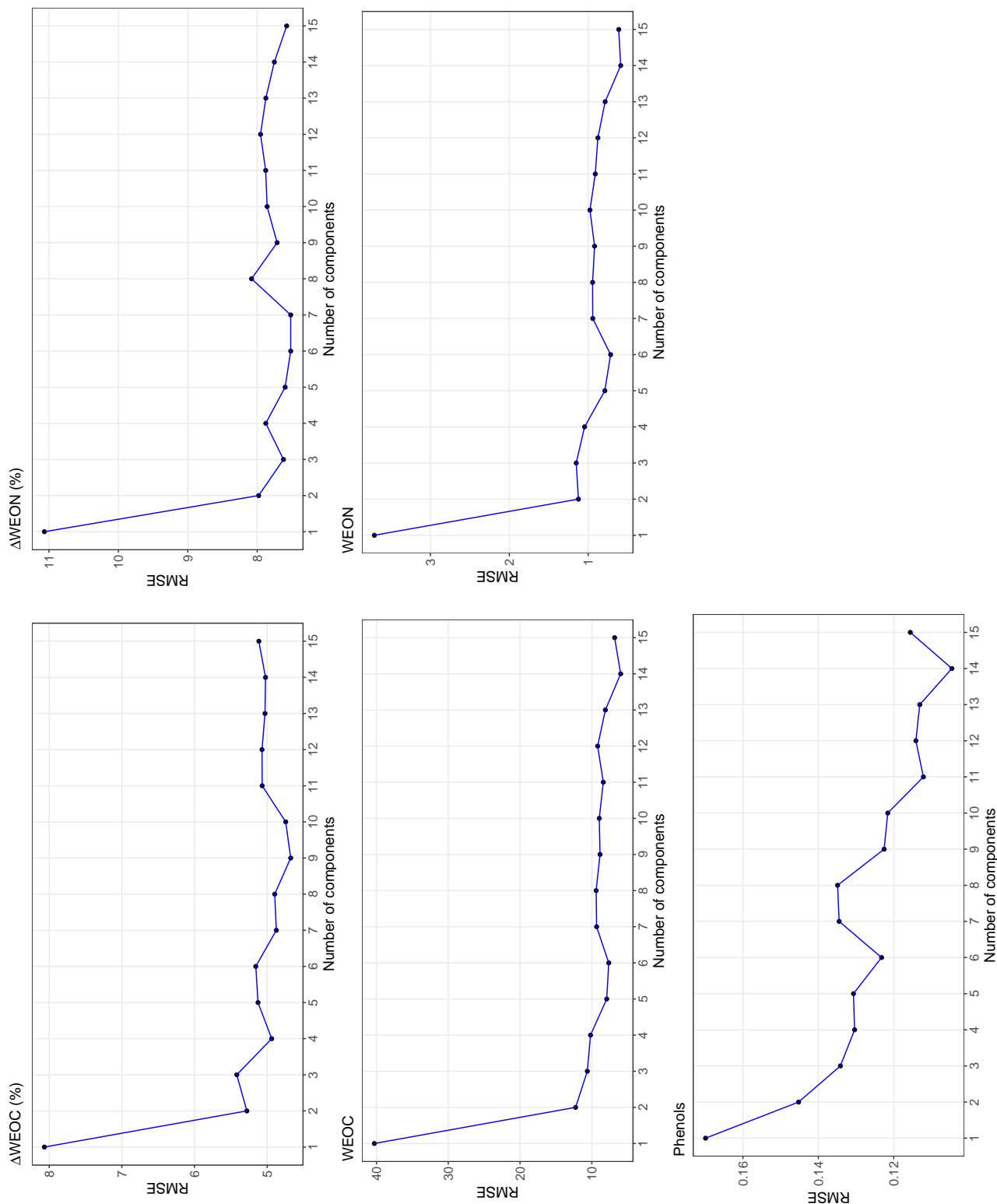
**Table C.1:** Summary statistics for the RF regression between actual and estimated WEOC and WEON utilization ( $\Delta$ WEOC and  $\Delta$ WEON), and pre- and post-incubation WEOC and WEON using six PARAFAC components.

	$\Delta$ WEOC	$\Delta$ WEON	WEOC	WEON
Correlation coefficient	0.90	0.78	0.98	0.98
Bias	0.066	-0.007	0.264	-0.001
Mean squared logarithmic error	0.03	0.02	1.5	NaN
Ratio of performance to deviation	2.16	1.33	5.56	5.38
$R^2$	0.81	0.61	0.97	0.97
Standard deviation of residuals	5.16	7.52	7.64	0.72
Mean absolute error	3.45	5.26	4.19	0.38
Root mean squared error	5.16	7.52	7.64	0.72
Relative absolute error (%)	37.69	58.81	13.13	13.87
Root relative squared error (%)	43.40	62.13	17.43	17.75
Total number of instances	107	107	214	214

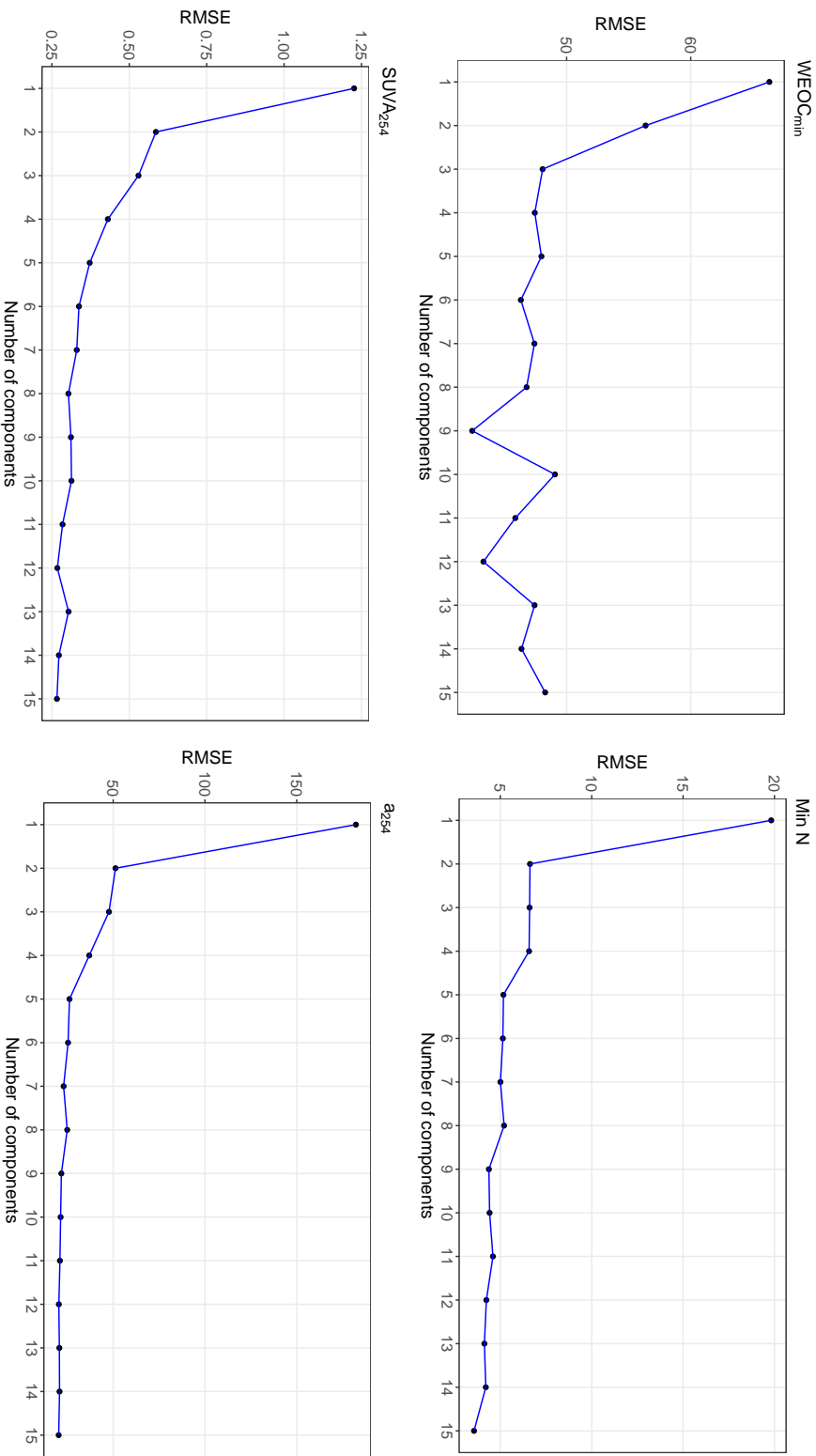


**Figure C.8:** Root mean square error (RMSE) of the RF algorithm in estimating WEOC mineralization (WEOC<sub>min</sub>) and mineral N (Min N) production over 14 days and WEOC pre- and post-incubation optical properties (SUV<sub>A254</sub> and a<sub>254</sub>) using PARAFAC components.

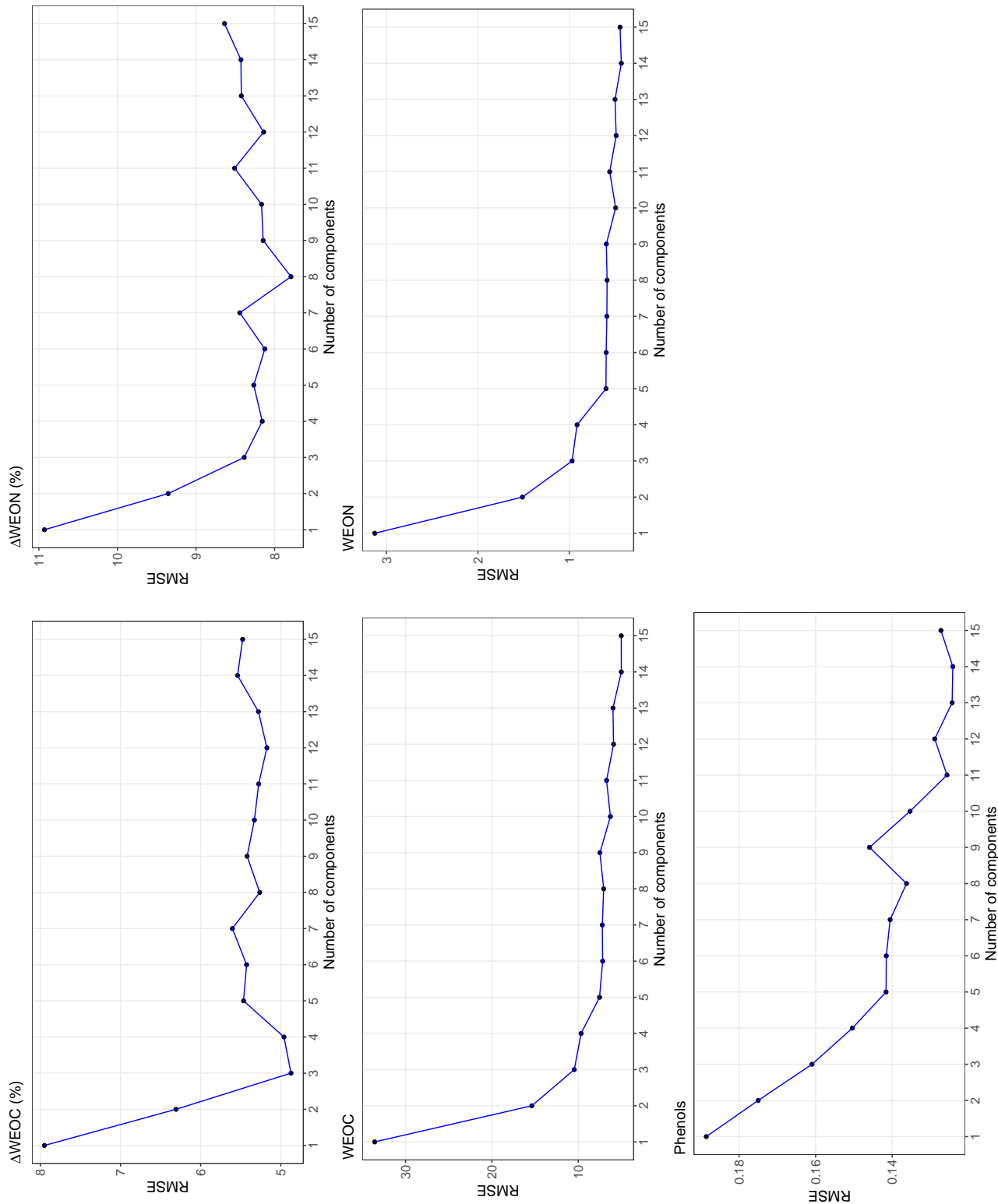




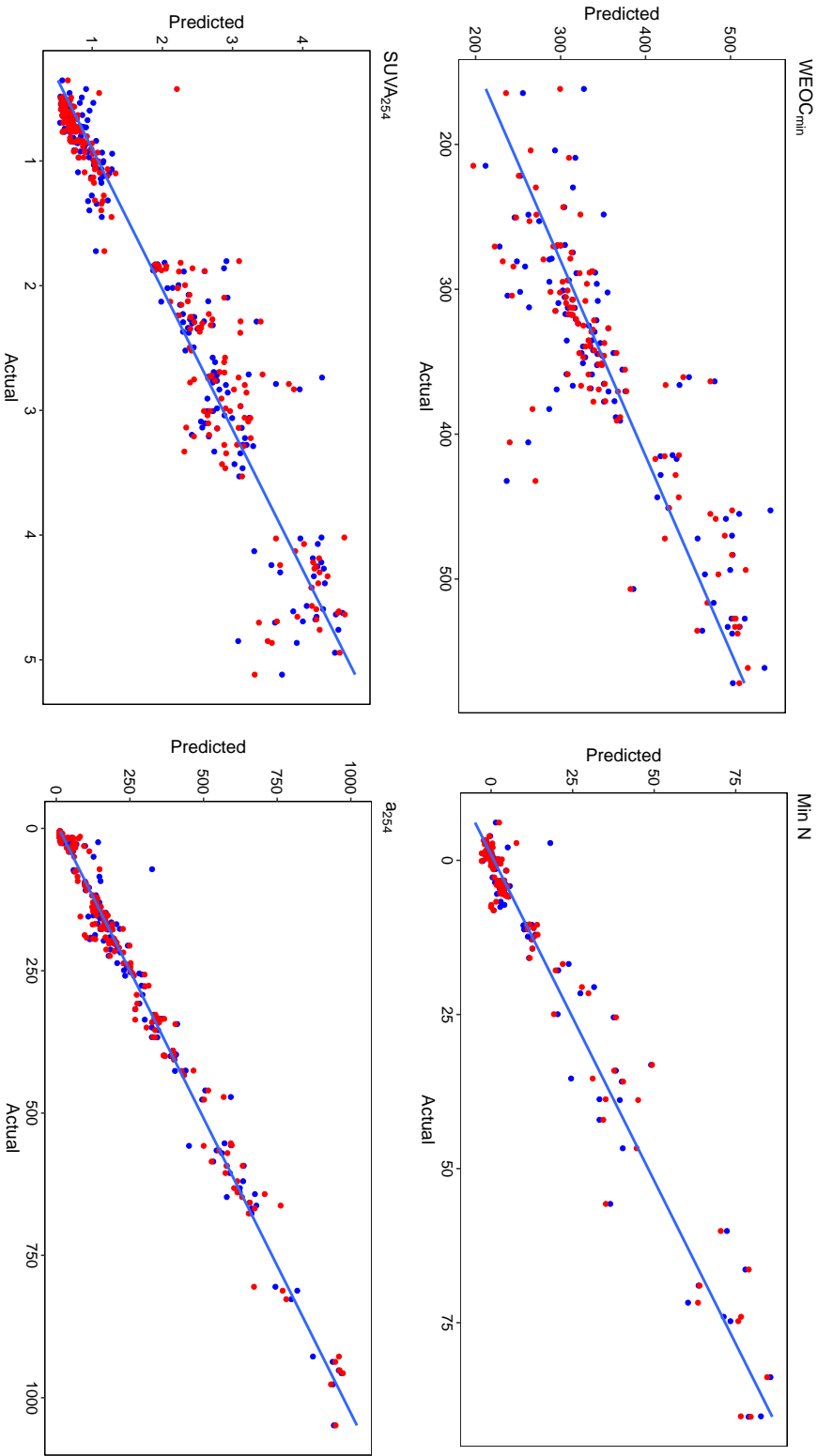
**Figure C.9:** Root mean square error (RMSE) of the RF algorithm in estimating WEOC and WEON utilization ( $\Delta WEOC$  and  $\Delta WEOC$ ) over 14 days, and pre- and post-incubation WEOC, WEON and phenols concentrations using PARAFAC components.



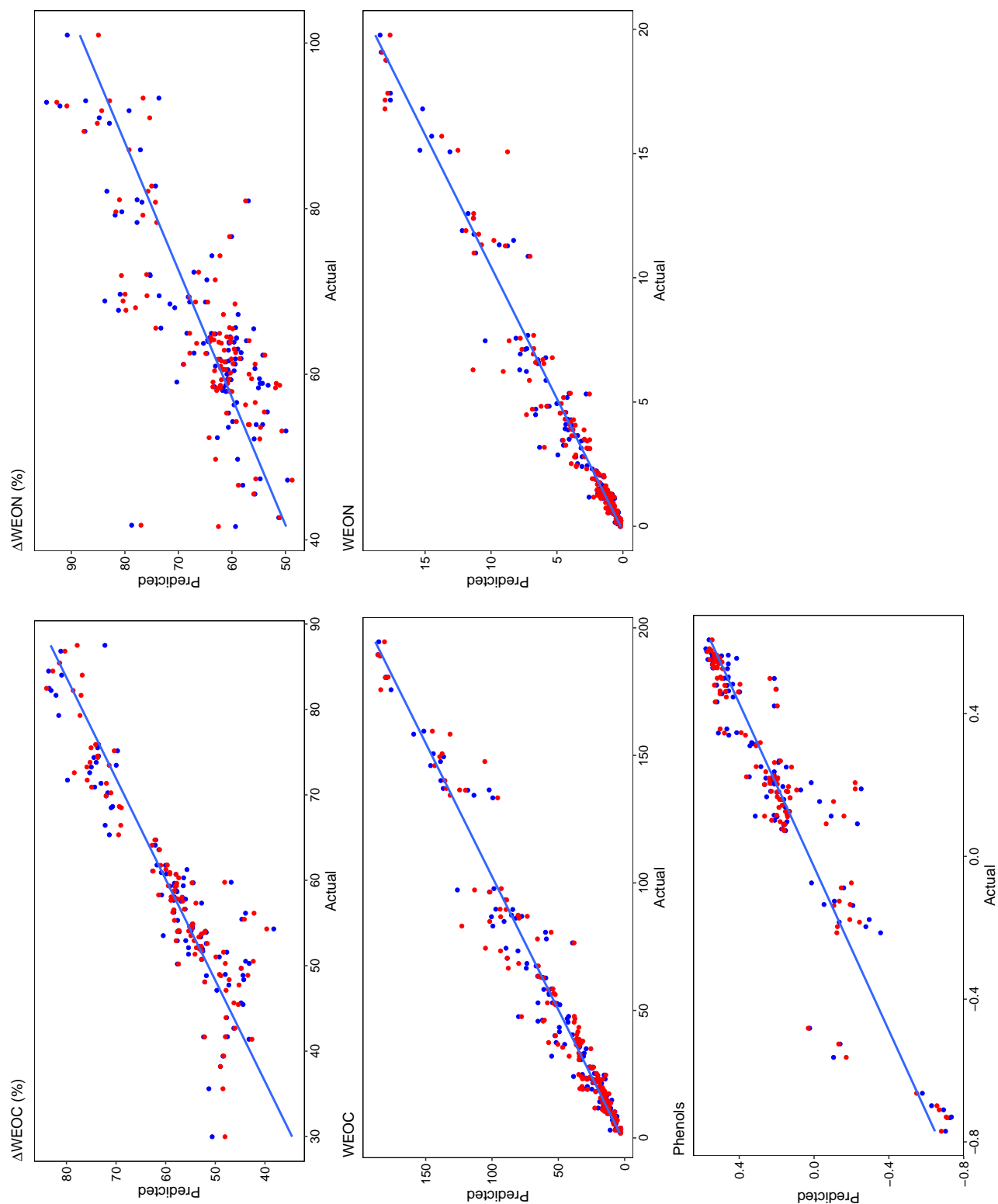
**Figure C.10:** Root mean square error (RMSE) of the LWL-GPD algorithm in estimating WEOC mineralization (WEOC<sub>min</sub>) and mineral N (Min N) production over 14 days and WEOM pre- and post-incubation optical properties (SUV<sub>A254</sub> and a<sub>254</sub>) using PARAFAC components.



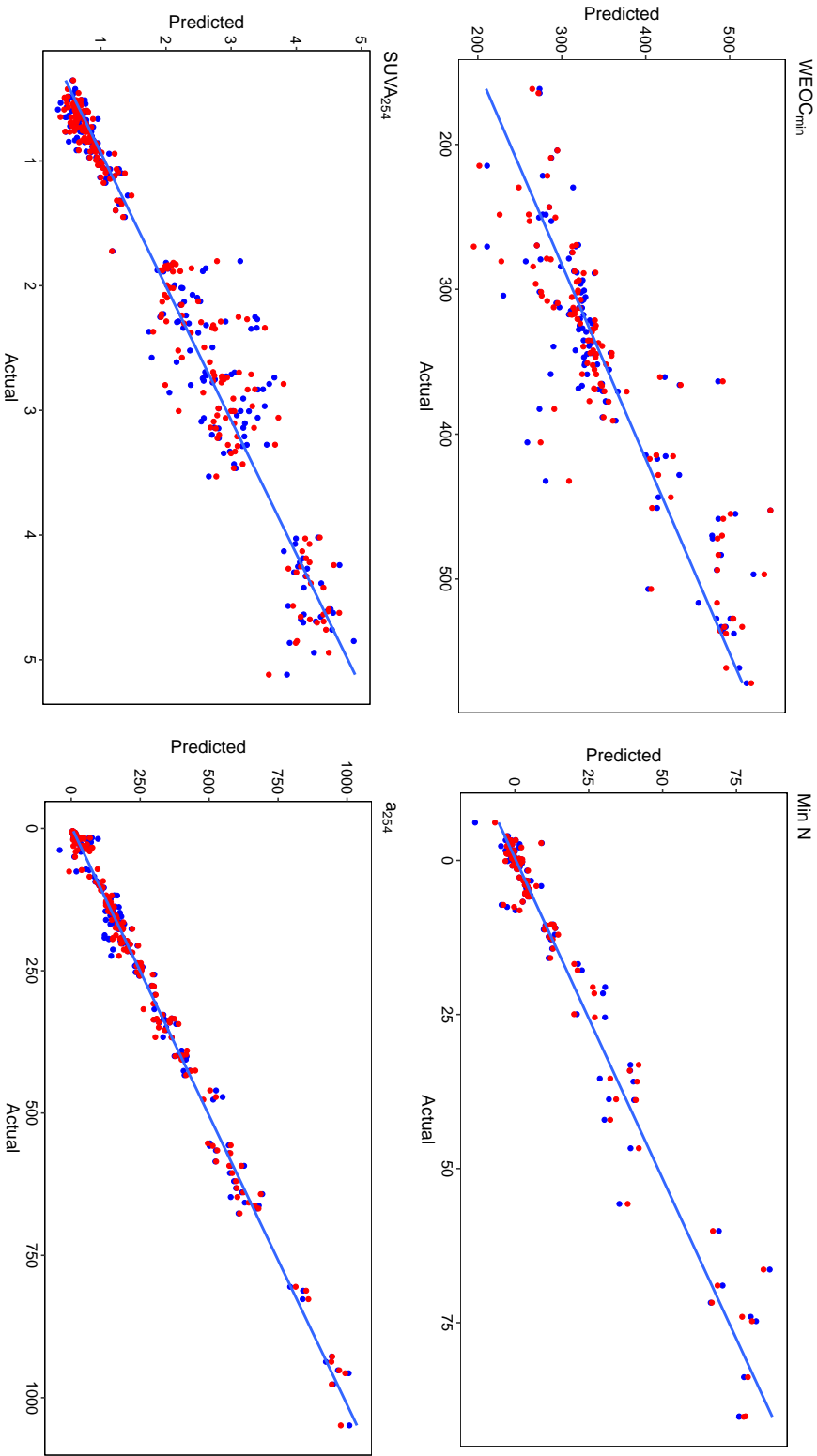
**Figure C.11:** Root mean square error (RMSE) of the LWL-GPD algorithm in estimating WEOC and WEON utilization ( $\Delta$ WEOC and  $\Delta$ WEON) over 14 days, and pre- and post-incubation WEOC, WEON and phenols concentrations using PARAFAC components.



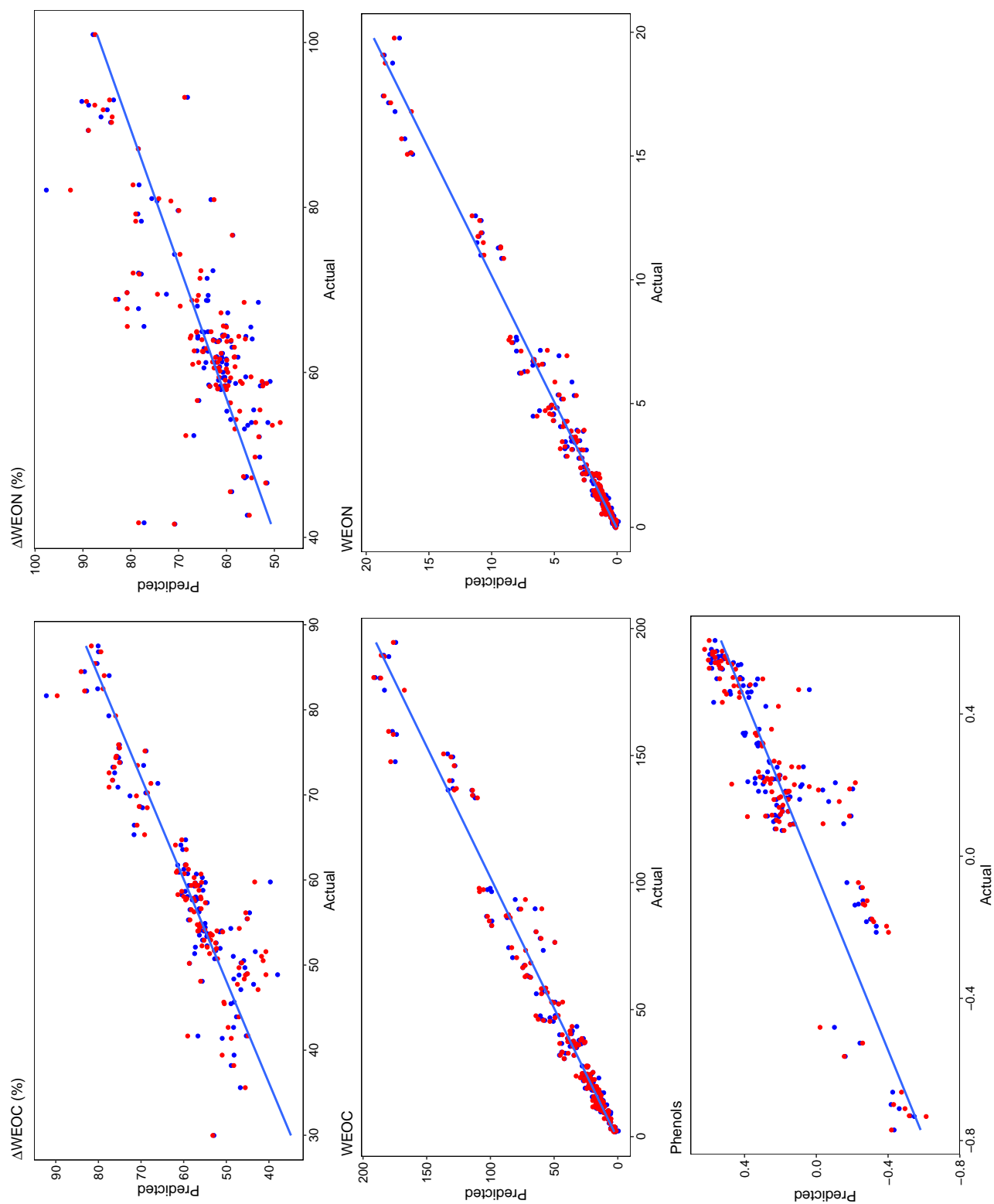
**Figure C.12:** Actual vs. predicted patterns of WEOC mineralization ( $WEOC_{min}$ ), mineral N (Min N) production, pre- and post-incubation WEOC optical properties ( $SUVA_{254}$  and  $a_{254}$ ) and phenolic content determined by six (blue dots) or nine (red dots) PARAFAC components using the RF algorithm.



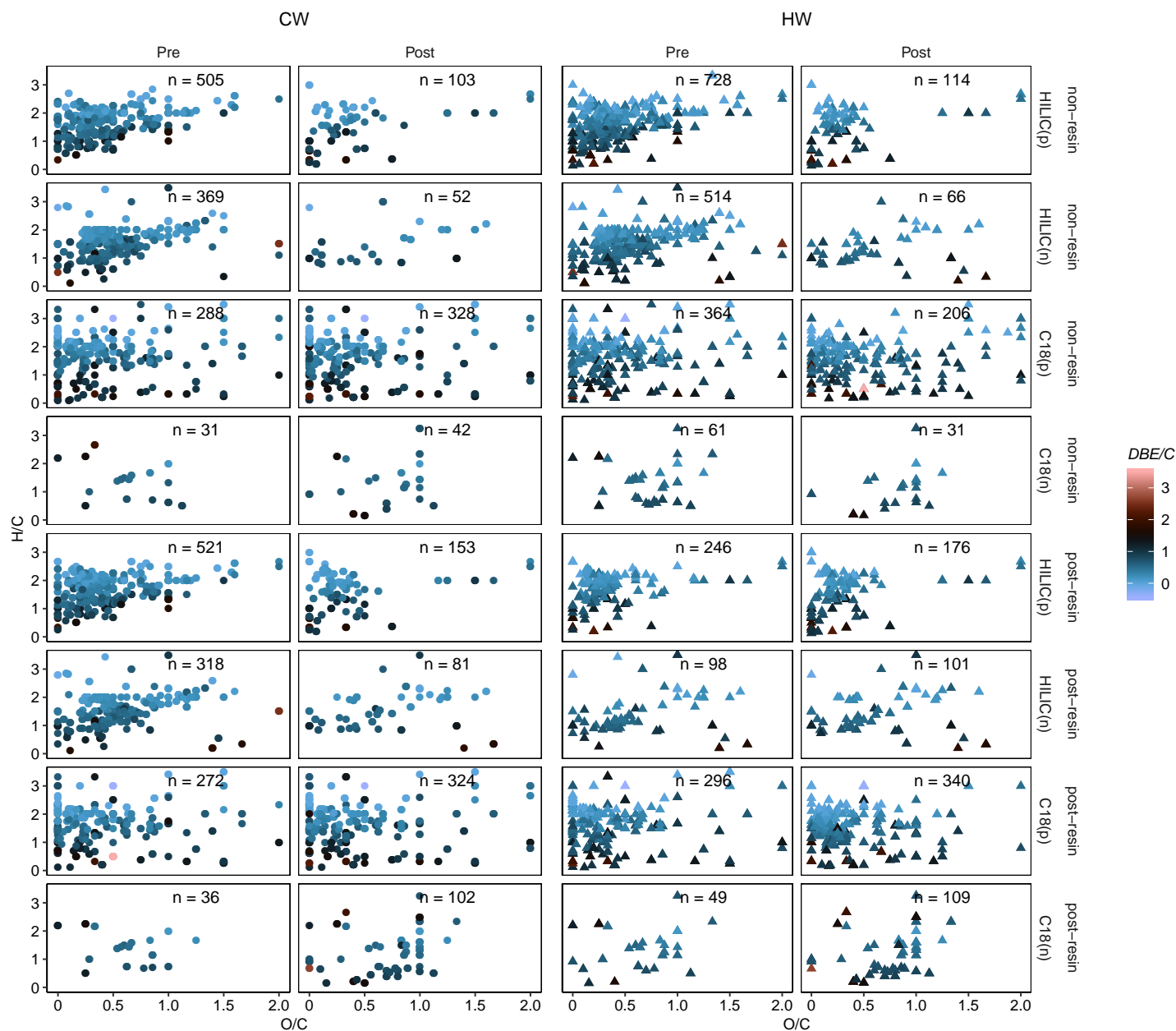
**Figure C.13:** Actual vs. predicted patterns of WEOC and WEON utilization ( $\Delta\text{WEOC}$  and  $\Delta\text{WEON}$ ) and pre- and post-incubation WEOC and WEON concentrations determined by six (blue dots) or nine (red dots) PARAFAC components using the RF algorithm.



**Figure C.14:** Actual vs. predicted patterns of WEOC mineralization ( $WEOC_{min}$ ), mineral N (Min N) production, pre- and post-incubation WEOC optical properties ( $SUVA_{254}$  and  $a_{254}$ ) and phenolic content determined by six (blue dots) or nine (red dots) PARAFAC components using the LWL-GPD algorithm.

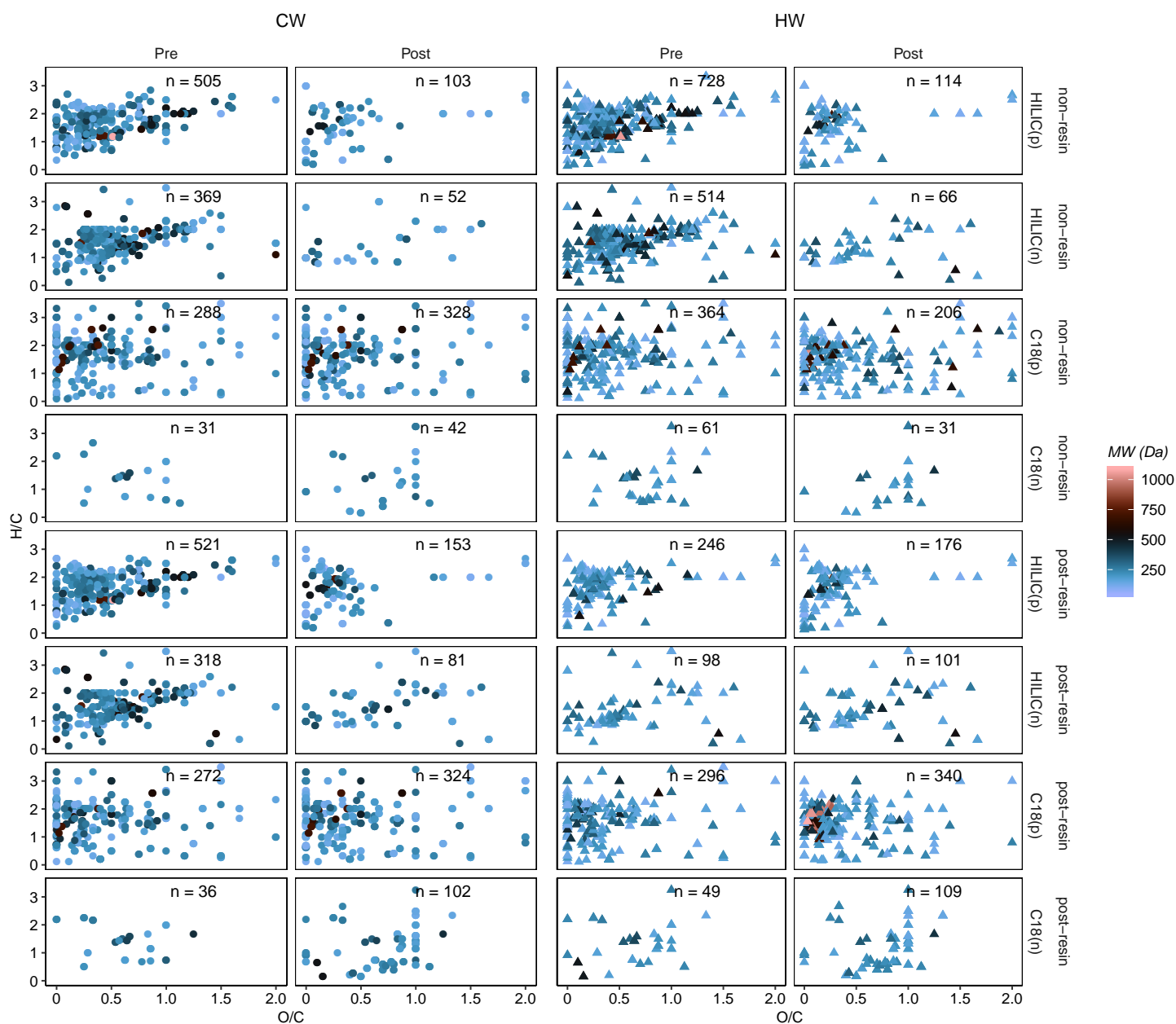


**Figure C.15:** Actual vs. predicted patterns of WEOC and WEON utilization ( $\Delta\text{WEOC}$  and  $\Delta\text{WEON}$ ) and pre- and post-incubation WEOC and WEON concentrations determined by six (blue dots) or nine (red dots) PARAFAC components using the LWL-GPD algorithm.

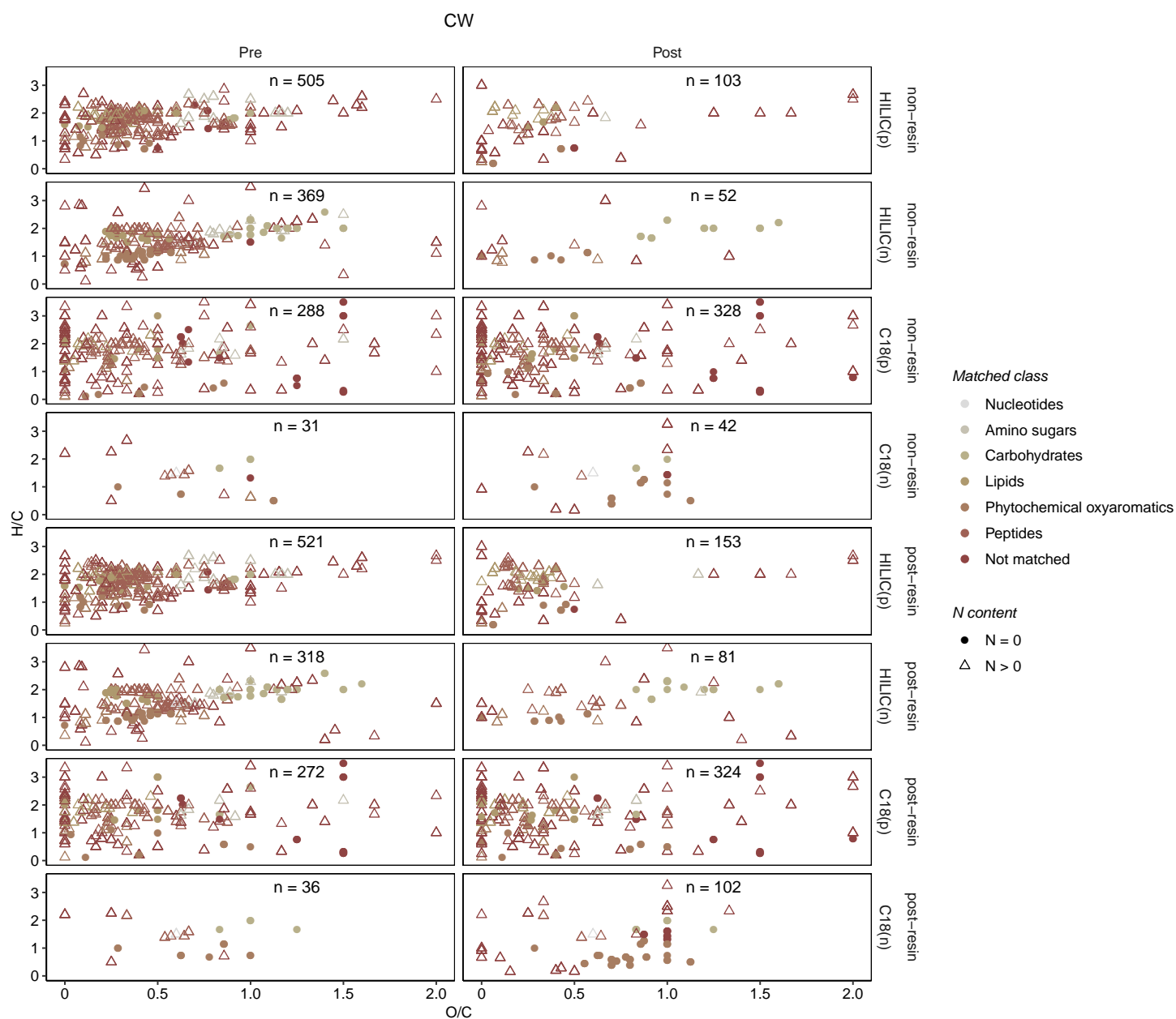


**Figure C.16:** van Krevelen diagrams of non- and post-resin treated, cold (CW) and hot water (HW) extracts before and after the 14-day incubation, plotted against DBE/C and relative to their chromatography/ionization mode of detection.  $n$  = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes.)

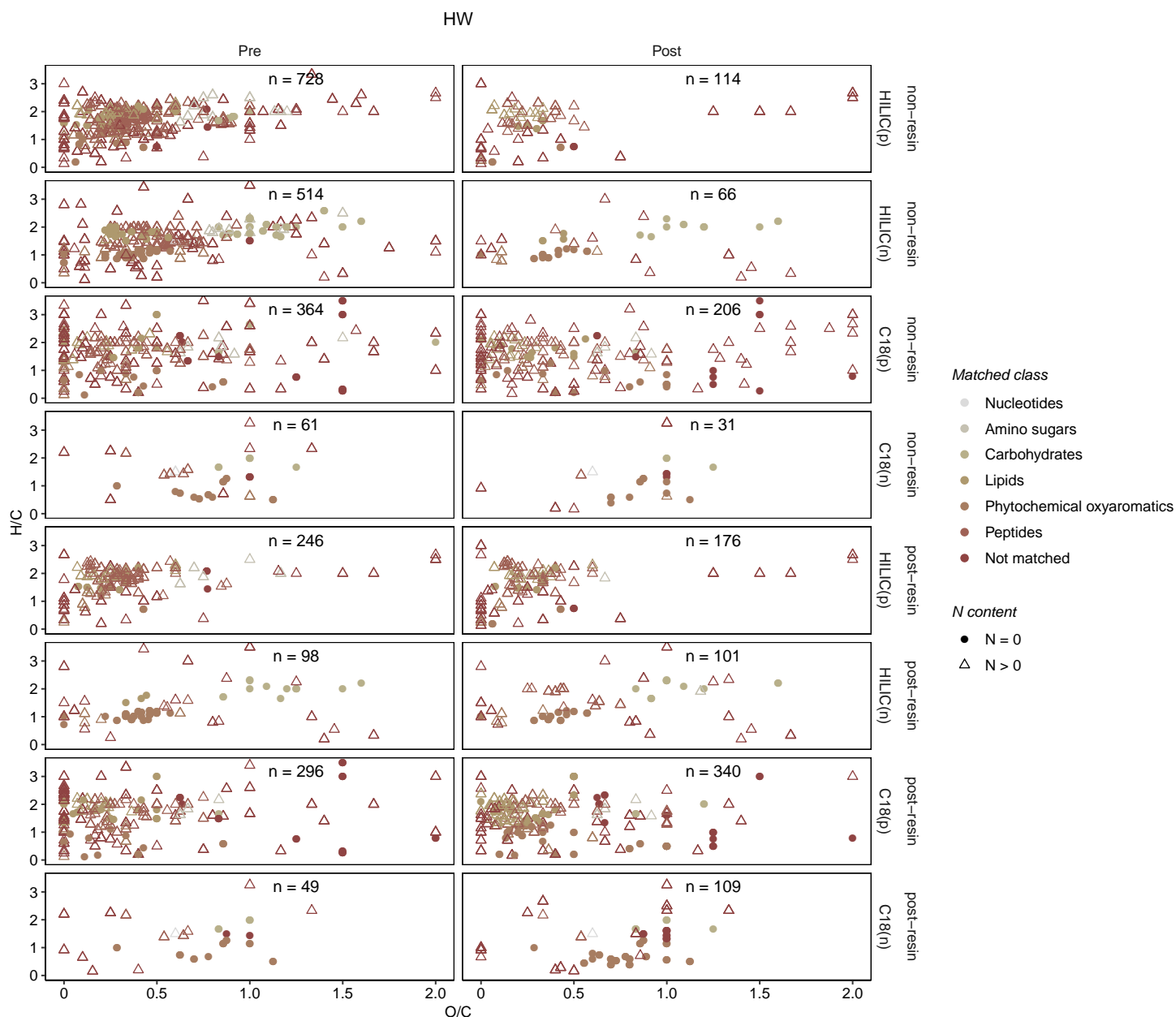




**Figure C.17:** van Krevelen diagrams of non- and post-resin treated, cold (CW) and hot water (HW) extracts before and after the 14-day incubation, plotted against the molecular weight (MW) and relative to their chromatography/ionization mode of detection. n = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes.)



**Figure C.18:** van Krevelen diagrams of compounds classified according to Rivas-Ubach et al. (2018) in non- and post-resin treated, cold water (CW) extracts before and after the 14-day incubation, relative to their chromatography/ionization mode of detection and presence or absence of nitrogen in their structure. n = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes.)



**Figure C.19:** van Krevelen diagrams of compounds classified according to Rivas-Ubach et al. (2018) in non- and post-resin treated, hot water (HW) extracts before and after the 14-day incubation, relative to their chromatography/ionization mode of detection and presence or absence of nitrogen in their structure. n = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes.)

**Table C.2:** Mean scores of PARAFAC components (C1–C5) and weighted average of molecular determinants of pre-incubation, non-resin treated cold (CW) and hot water (HW) extracts of different land uses utilized for PCA analysis.

Pre-Incubation																		
Extraction	Resin treatment	Land use	Replicate	C1	C2	C3	C4	C5	MW	#C	#N	#P	O/C	H/C	Al <sub>mod</sub>	DBE/C	DBE-O	
	Fallow	1	1	1.02	0.19	0.40	-0.02	0.21	238.40	8.07	2.48	0.24	0.72	1.73	0.48	0.53	-1.33	
			2	0.77	0.16	0.49	0.03	0.24	244.19	8.29	2.56	0.23	0.71	1.74	0.39	0.52	-1.42	
			3	0.90	0.19	0.58	0.04	0.29	246.45	8.45	2.58	0.25	0.69	1.72	0.40	0.52	-1.29	
		1	1	0.97	0.20	0.55	0.03	0.29	275.47	9.27	2.87	0.16	0.65	1.77	-0.02	0.44	-2.17	
			2	1.00	0.22	0.55	0.03	0.31	280.33	9.49	2.82	0.15	0.66	1.76	-0.03	0.43	-2.39	
			3	1.00	0.22	0.57	0.03	0.31	282.15	9.56	2.92	0.18	0.66	1.75	-0.01	0.45	-2.11	
		1	1	1.37	0.28	0.80	0.01	0.37	255.45	8.96	2.83	0.18	0.58	1.78	0.13	0.45	-1.21	
			2	1.58	0.31	0.91	0.03	0.44	246.98	8.70	2.66	0.17	0.56	1.82	0.05	0.42	-1.36	
			3	1.52	0.30	0.93	0.00	0.41	252.46	8.90	2.66	0.18	0.58	1.78	0.14	0.44	-1.38	
CW	Pasture	1	1	2.57	0.56	0.90	0.17	0.71	237.10	8.53	2.95	0.11	0.53	1.65	0.20	0.54	-0.21	
			2	2.95	0.61	1.07	0.14	0.80	232.99	8.39	2.98	0.11	0.52	1.65	0.15	0.55	-0.03	
			3	3.61	0.73	1.34	0.17	0.99	231.65	8.41	2.93	0.11	0.52	1.62	0.14	0.55	0.05	
	Fallow	1	1	3.78	0.81	1.23	0.16	0.93	220.15	8.48	2.53	0.10	0.47	1.63	-0.10	0.50	0.02	
			2	3.84	0.82	1.25	0.15	0.96	220.78	8.51	2.55	0.09	0.47	1.62	-0.10	0.51	0.06	
			3	3.44	0.75	1.11	0.15	0.87	221.49	8.57	2.53	0.10	0.46	1.64	-0.12	0.49	-0.03	
	HW	No-tillage	1	1	4.57	0.90	1.65	0.07	1.02	234.95	8.87	2.68	0.14	0.50	1.62	-0.01	0.51	-0.14
				2	5.45	1.08	1.91	0.12	1.19	232.48	8.79	2.59	0.14	0.50	1.63	-0.03	0.50	-0.20
				3	5.05	0.97	1.93	0.03	1.12	233.28	8.86	2.66	0.13	0.49	1.63	-0.03	0.51	-0.13
Pasture	2	1	5.45	1.08	1.91	0.12	1.19	232.48	8.79	2.59	0.14	0.50	1.63	-0.03	0.50	-0.20		
		2	5.45	1.08	1.91	0.12	1.19	232.48	8.79	2.59	0.14	0.50	1.63	-0.03	0.50	-0.20		
		3	5.05	0.97	1.93	0.03	1.12	233.28	8.86	2.66	0.13	0.49	1.63	-0.03	0.51	-0.13		

Abbreviations: MW, molecular weight (Da); #C, #N and #P, number of carbon, nitrogen, phosphorus atoms; O/C and H/C, oxygen and hydrogen to carbon ratios, respectively; Al<sub>mod</sub>, modified aromaticity index; DBE/C and DBE-O, double bond equivalents per unit carbon and minus oxygen, respectively.

**Table C.3:** Mean scores of PARAFAC components (C1–C5) and weighted average of molecular determinants of pre-incubation, post-resin treated cold (CW) and hot water (HW) extracts of different land uses utilized for PCA analysis.

Pre-Incubation																	
Extraction	Resin treatment	Land use	Replicate	C1	C2	C3	C4	C5	MW	#C	#N	#P	O/C	H/C	AI <sub>mod</sub>	DBE/C	DBE-O
CW	post-	Fallow	1	0.59	0.23	0.11	0.16	0.10	169.37	7.16	2.25	0.09	0.21	2.34	-0.30	0.16	-0.50
			2	0.52	0.22	0.10	0.09	0.10	168.16	7.15	2.28	0.09	0.20	2.35	-0.30	0.16	-0.43
			3	0.58	0.25	0.09	0.09	0.11	156.17	6.87	2.20	0.07	0.16	2.41	-0.34	0.13	-0.35
		No-tillage	1	0.85	0.30	0.11	0.12	0.17	186.93	7.56	2.31	0.10	0.29	2.23	-0.26	0.21	-0.77
			2	0.84	0.31	0.11	0.08	0.18	194.73	7.80	2.38	0.10	0.30	2.19	-0.24	0.23	-0.79
			3	0.79	0.30	0.11	0.08	0.18	184.44	7.53	2.28	0.10	0.28	2.24	-0.25	0.20	-0.74
		Pasture	1	1.04	0.38	0.09	0.08	0.19	208.95	8.10	2.51	0.08	0.36	2.12	-0.14	0.26	-0.95
			2	1.00	0.39	0.10	0.11	0.19	211.15	8.06	2.48	0.09	0.39	2.07	-0.13	0.29	-0.98
			3	1.04	0.35	0.11	0.01	0.16	216.95	8.28	2.60	0.09	0.39	2.06	-0.09	0.29	-0.90
HW	post-	Fallow	2	2.31	1.04	0.27	0.65	0.28	139.74	6.77	2.10	0.04	0.11	2.46	-0.30	0.11	0.15
			3	2.45	1.11	0.30	0.67	0.30	136.21	6.60	2.09	0.04	0.10	2.48	-0.32	0.10	0.14
			1	3.81	1.54	0.48	1.18	0.43	130.84	6.16	2.06	0.05	0.12	2.46	-0.34	0.13	0.11
		No-tillage	2	4.21	1.65	0.54	1.26	0.50	131.51	6.14	2.08	0.06	0.12	2.44	-0.34	0.14	0.15
			3	3.66	1.46	0.46	1.10	0.43	130.60	6.11	2.07	0.06	0.12	2.45	-0.35	0.14	0.14
			1	5.17	1.93	0.66	2.17	0.39	127.21	5.98	2.03	0.05	0.14	2.47	-0.30	0.13	0.07
		Pasture	2	5.72	2.22	0.74	2.43	0.47	129.12	5.91	2.06	0.06	0.18	2.43	-0.24	0.17	0.10
			3	5.02	1.62	0.75	2.22	0.43	130.17	5.90	2.09	0.05	0.19	2.40	-0.25	0.19	0.11

Abbreviations: MW, molecular weight (Da); #C, #N and #P, number of carbon, nitrogen, phosphorus atoms; O/C and H/C, oxygen and hydrogen to carbon ratios, respectively; AI<sub>mod</sub>, modified aromaticity index; DBE/C and DBE-O, double bond equivalents per unit carbon and minus oxygen, respectively.

**Table C.4:** Mean scores of PARAFAC components (C1–C5) and weighted average of molecular determinants of post-incubation, non-resin treated cold (CW) and hot water (HW) extracts of different land uses utilized for PCA analysis.

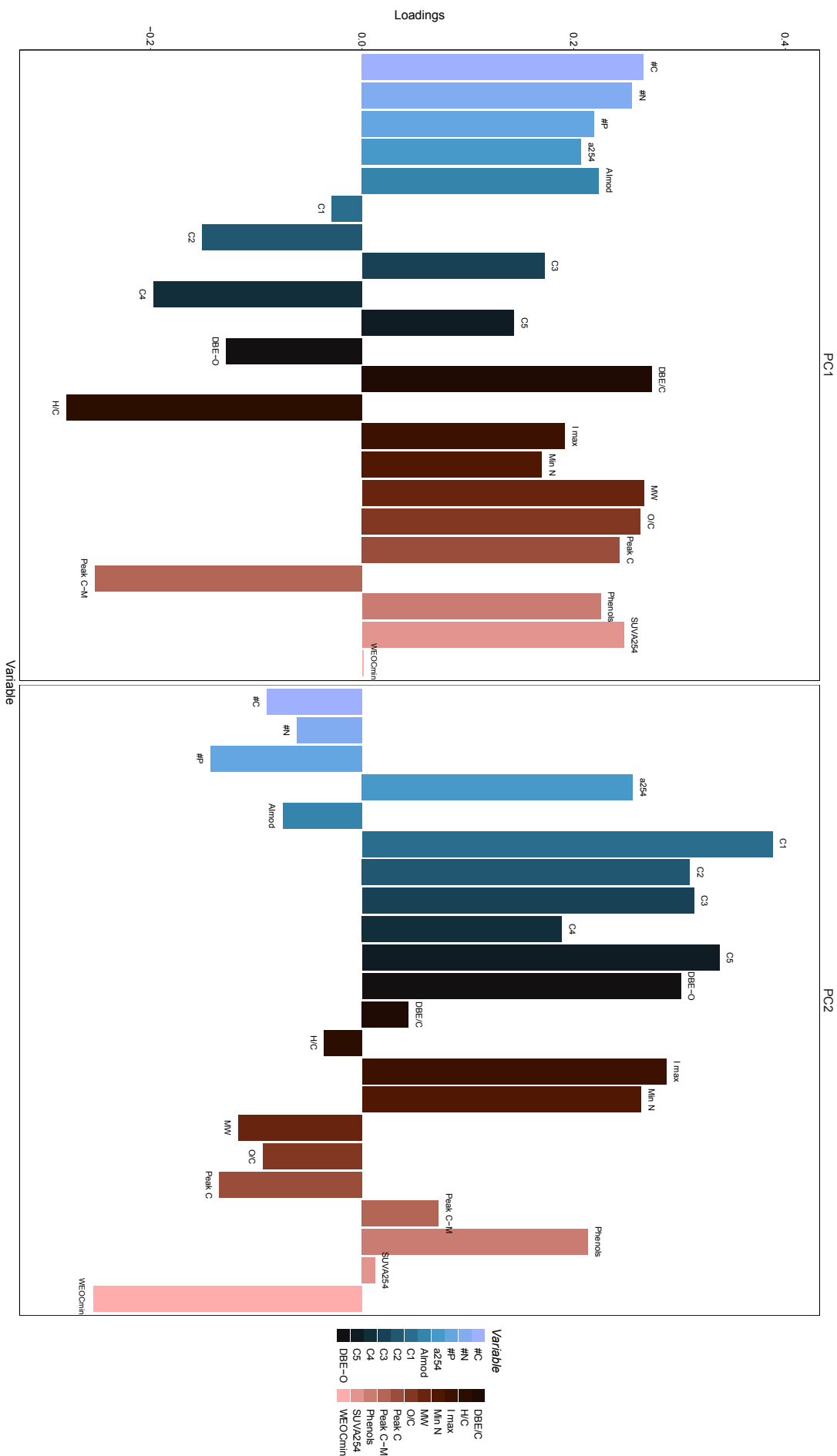
Post-Incubation																	
Extraction	Resin treatment	Land use	Replicate	C1	C2	C3	C4	C5	MW	#C	#N	#P	O/C	H/C	Al <sub>mod</sub>	DBE/C	DBE-O
CW	non-	Fallow	1	0.62	0.13	0.35	0.04	0.20	177.95	5.60	2.38	0.29	0.52	1.55	0.35	0.74	1.60
			2	0.66	0.13	0.40	0.03	0.21	181.72	5.62	2.46	0.29	0.56	1.54	0.27	0.75	1.42
			3	0.77	0.16	0.48	0.04	0.26	182.88	5.60	2.51	0.29	0.55	1.51	0.31	0.78	1.66
		No-tillage	1	0.75	0.15	0.43	0.04	0.24	168.20	4.45	2.04	0.23	1.06	1.79	0.80	0.71	-0.45
			2	0.78	0.16	0.45	0.04	0.26	173.80	4.72	2.05	0.26	0.96	1.70	0.61	0.73	-0.09
			3	0.78	0.16	0.45	0.04	0.25	173.44	4.72	1.97	0.27	0.95	1.69	0.61	0.72	-0.12
	Pasture	1	1.09	0.21	0.67	0.05	0.35	161.81	4.77	2.15	0.21	0.83	1.74	0.83	0.72	0.52	
		2	1.27	0.25	0.75	0.06	0.41	161.35	4.62	2.27	0.21	0.87	1.76	0.80	0.74	0.37	
		3	1.18	0.23	0.79	0.06	0.39	160.56	4.59	2.13	0.21	0.88	1.79	0.87	0.71	0.27	
		Fallow	1	2.06	0.47	0.73	0.15	0.60	152.48	3.55	2.02	0.12	1.23	1.84	1.06	0.78	-0.74
			2	2.36	0.52	0.89	0.16	0.70	149.82	3.58	1.98	0.13	1.10	1.80	0.92	0.81	-0.19
			3	2.76	0.59	1.06	0.17	0.81	151.50	3.51	2.02	0.16	1.17	1.83	0.98	0.79	-0.63
HW	non-	No-tillage	1	2.84	0.64	0.99	0.18	0.80	176.95	4.96	1.79	0.14	1.05	1.80	0.87	0.68	-0.71
			2	2.99	0.68	1.03	0.20	0.85	181.81	5.17	1.77	0.16	0.99	1.76	0.71	0.68	-0.49
			3	2.72	0.62	0.91	0.18	0.79	177.13	5.00	1.80	0.16	1.01	1.75	0.74	0.71	-0.39
		Pasture	1	3.59	0.78	1.46	0.21	1.04	161.05	4.47	1.92	0.25	0.86	1.74	0.62	0.75	0.34
			2	4.23	0.93	1.65	0.24	1.19	172.94	4.84	1.87	0.26	0.80	1.65	0.29	0.74	0.45
			3	3.80	0.82	1.65	0.21	1.15	165.40	4.49	1.81	0.24	0.92	1.79	0.49	0.70	-0.03

Abbreviations: MW, molecular weight (Da); #C, #N and #P, number of carbon, nitrogen, phosphorus atoms; O/C and H/C, oxygen and hydrogen to carbon ratios, respectively; Al<sub>mod</sub>, modified aromaticity index; DBE/C and DBE-O, double bond equivalents per unit carbon and minus oxygen, respectively.

**Table C.5:** Mean scores of PARAFAC components (C1–C5) and weighted average of molecular determinants of post-incubation, post-resin treated cold (CW) and hot water (HW) extracts of different land uses utilized for PCA analysis.

Post-Incubation																	
Extraction	Resin treatment	Land use	Replicate	C1	C2	C3	C4	C5	MW	#C	#N	#P	O/C	H/C	AI <sub>mod</sub>	DBE/C	DBE-O
CW	post-	Fallow	1	0.45	0.21	0.08	0.10	0.08	159.38	4.97	2.26	0.18	0.71	1.88	0.37	0.61	0.07
			2	0.47	0.22	0.07	0.11	0.08	156.92	5.23	2.33	0.17	0.57	1.93	0.32	0.56	0.39
			3	0.49	0.23	0.08	0.11	0.08	152.59	5.11	2.21	0.13	0.61	2.01	0.23	0.52	-0.02
		No-tillage	1	0.67	0.29	0.09	0.13	0.11	139.53	5.78	2.13	0.11	0.25	2.25	-0.24	0.30	0.38
			2	0.70	0.30	0.09	0.13	0.12	165.42	5.78	2.39	0.24	0.43	1.86	-0.02	0.57	0.83
			3	0.63	0.28	0.09	0.12	0.12	143.11	5.87	2.17	0.12	0.25	2.25	-0.25	0.30	0.34
		Pasture	1	0.65	0.26	0.08	0.12	0.12	138.00	5.75	2.10	0.13	0.27	2.21	-0.12	0.33	0.48
			2	0.74	0.30	0.09	0.14	0.13	153.28	5.70	2.23	0.22	0.38	1.89	0.06	0.54	0.95
			3	0.73	0.29	0.10	0.14	0.13	155.12	5.54	2.29	0.23	0.44	1.79	0.13	0.62	1.11
HW	post-	Fallow	1	1.83	0.88	0.16	0.45	0.25	186.72	6.38	1.89	0.12	0.90	1.74	0.61	0.69	-0.22
			2	2.08	0.97	0.19	0.51	0.26	187.74	6.37	1.89	0.12	0.90	1.74	0.56	0.70	-0.30
			3	2.12	0.99	0.20	0.53	0.30	185.22	6.21	1.86	0.14	0.90	1.70	0.58	0.71	-0.15
		No-tillage	1	3.12	1.38	0.31	0.78	0.45	177.83	5.84	1.92	0.16	0.95	1.71	0.82	0.74	-0.03
			2	3.33	1.44	0.34	0.84	0.49	173.46	5.49	1.91	0.13	0.96	1.73	0.86	0.74	-0.12
			3	2.94	1.30	0.28	0.75	0.40	174.93	5.60	1.93	0.14	0.93	1.72	0.83	0.74	-0.06
		Pasture	1	3.31	1.29	0.29	1.23	0.38	154.06	6.94	2.01	0.08	0.16	2.24	-0.32	0.25	0.63
			2	3.88	1.53	0.37	1.42	0.48	150.75	6.69	1.99	0.08	0.17	2.27	-0.39	0.24	0.46
			3	3.78	1.34	0.38	1.45	0.46	148.50	6.63	1.99	0.08	0.16	2.26	-0.39	0.24	0.48

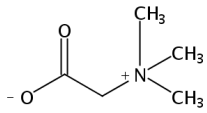
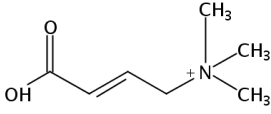
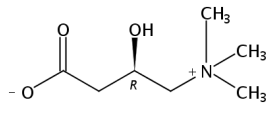
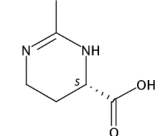
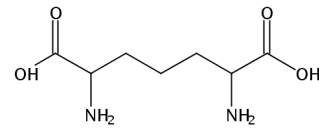
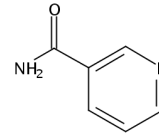
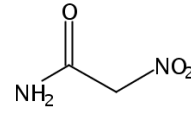
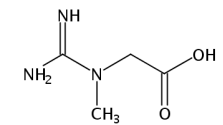
Abbreviations: MW, molecular weight (Da); #C, #N and #P, number of carbon, nitrogen, phosphorus atoms; O/C and H/C, oxygen and hydrogen to carbon ratios, respectively; AI<sub>mod</sub>, modified aromaticity index; DBE/C and DBE-O, double bond equivalents per unit carbon and minus oxygen, respectively.



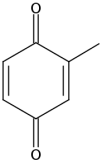
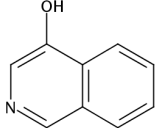
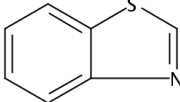
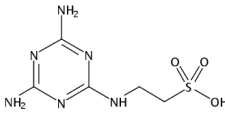
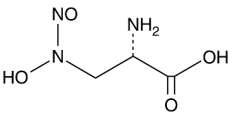
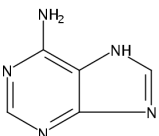
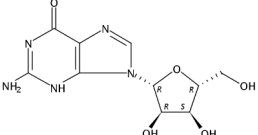
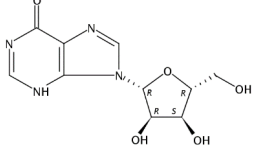
**Figure C.20:** Principal component analysis loadings of PARAFAC component scores (C1–C5), weighted molecular determinants, WEOCmn optical properties and biodegradability indicators for PC1 and PC2.



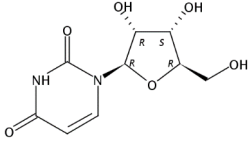
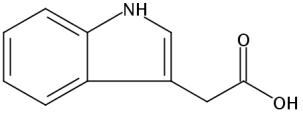
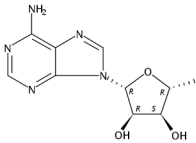
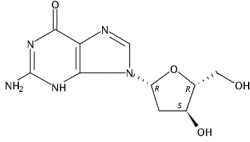
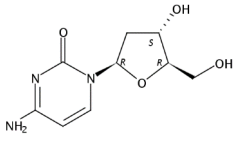
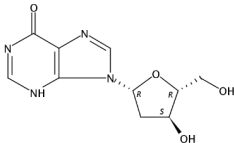
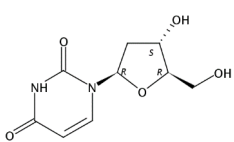
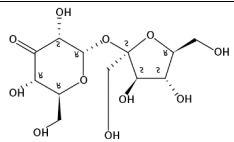
**Table C.6:** List of compounds discussed in Chapter 5, including the molecular formulae, monoisotopic masses, structures and chromatography/ionization modes of detection.

Mode	Compound name	Molecular formulae	Monoisotopic mass	Molecular structure
HILIC(p)	Betaine	C <sub>5</sub> H <sub>11</sub> NO <sub>2</sub>	117.079	
HILIC(p)	Crotono-betaine	C <sub>7</sub> H <sub>14</sub> NO <sub>2</sub>	143.095	
HILIC(p)	L(-)-Carnitine	C <sub>7</sub> H <sub>15</sub> NO <sub>3</sub>	161.105	
HILIC(p)	Ectoine	C <sub>6</sub> H <sub>10</sub> N <sub>2</sub> O <sub>2</sub>	142.074	
HILIC(p) and (n)	DL-α,ε-Diaminopimelic acid	C <sub>7</sub> H <sub>14</sub> N <sub>2</sub> O <sub>4</sub>	190.095	
HILIC(p)	Nicotinamide	C <sub>6</sub> H <sub>6</sub> N <sub>2</sub> O	122.048	
HILIC(p)	Nitroacetamide	C <sub>2</sub> H <sub>4</sub> N <sub>2</sub> O <sub>3</sub>	104.022	
HILIC(p)	Creatine	C <sub>4</sub> H <sub>9</sub> N <sub>3</sub> O <sub>2</sub>	131.069	

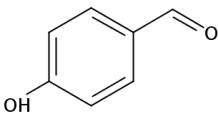
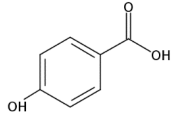
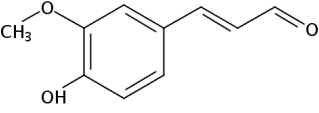
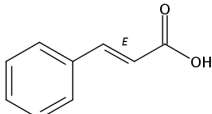
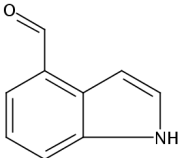
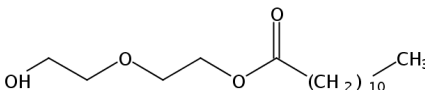
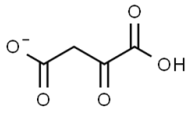
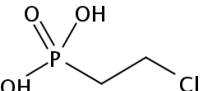
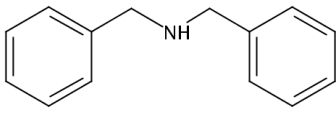
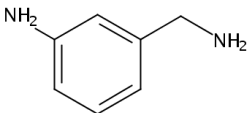
**Table C.6:** List of compounds discussed in Chapter 5, including the molecular formulae, monoisotopic masses, structures and chromatography/ionization modes of detection. (cont.)

HILIC(p)	2-Methyl-1,4-benzoquinone	$C_7H_6O_2$	122.037	
HILIC(p)	4-Isoquinolinol	$C_9H_7NO$	145.053	
HILIC and C18 (p)	1,3-Benzothiazole	$C_7H_5NS$	135.014	
HILIC(p)	2-[(4,6-diamino-1,3,5-triazin-2-yl)amino]ethanesulfonic acid	$C_5H_{10}N_6O_3S$	234.053	
HILIC and C18(n)	L-Alanosine	$C_3H_7N_3O_4$	149.044	
HILIC(p) and (n), C18(p)	Adenine	$C_5H_5N_5$	135.054	
HILIC(p) and (n)	Guanosine	$C_{10}H_{13}N_5O_5$	283.092	
HILIC(p) and (n)	Inosine	$C_{10}H_{12}N_4O_5$	268.081	

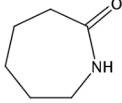
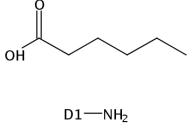
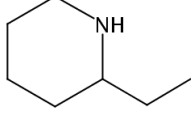
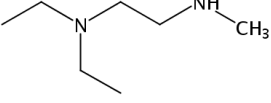
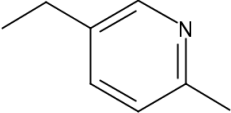
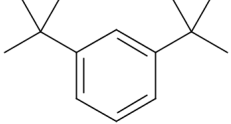
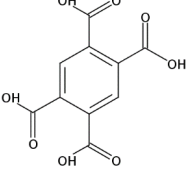
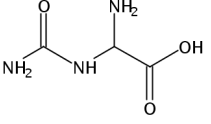
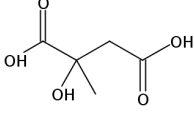
**Table C.6:** List of compounds discussed in Chapter 5, including the molecular formulae, monoisotopic masses, structures and chromatography/ionization modes of detection. (cont.)

HILIC(n)	Uridine	$C_9H_{12}N_2O_6$	244.07	
HILIC(n)	Indole-3-acetic acid	$C_{10}H_9NO_2$	175.063	
HILIC(p)	5'-Deoxyadenosine	$C_{10}H_{13}N_5O_3$	251.102	
HILIC(p)	2'-Deoxyguanosine	$C_{10}H_{13}N_5O_4$	267.097	
HILIC(p)	2'-Deoxycytidine	$C_9H_{13}N_3O_4$	227.091	
HILIC(n)	2'-Deoxyinosine	$C_{10}H_{12}N_4O_4$	252.086	
HILIC(n)	2'-Deoxyuridine	$C_9H_{12}N_2O_5$	228.075	
HILIC(n)	3'-Ketosucrose	$C_{12}H_{20}O_{11}$	340.101	

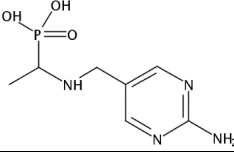
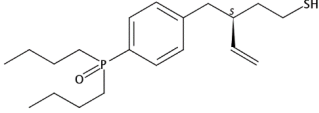
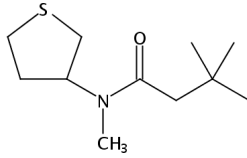
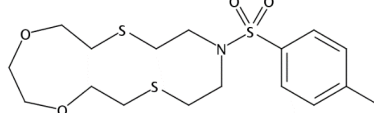
**Table C.6:** List of compounds discussed in Chapter 5, including the molecular formulae, monoisotopic masses, structures and chromatography/ionization modes of detection. (cont.)

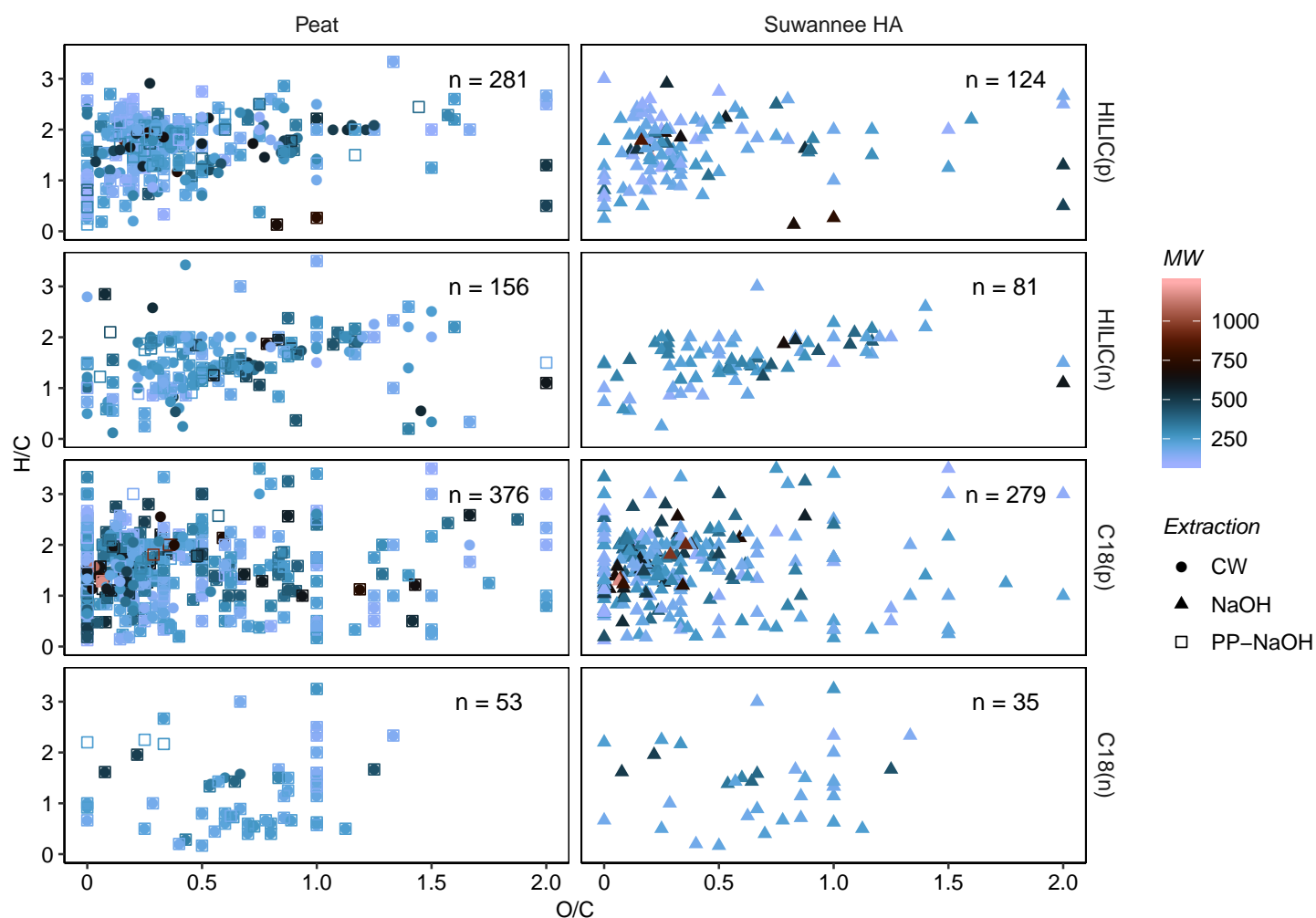
HILIC(n)	4-Hydroxybenzaldehyde	C <sub>7</sub> H <sub>6</sub> O <sub>2</sub>	122.037	
HILIC(n)	4-Hydroxybenzoic acid	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>	138.032	
HILIC(n)	4-hydroxy-3-methoxycinnamaldehyde	C <sub>10</sub> H <sub>10</sub> O <sub>3</sub>	178.063	
HILIC(n)	trans-Cinnamic acid	C <sub>9</sub> H <sub>8</sub> O <sub>2</sub>	148.052	
HILIC(n)	4-Indolecarbaldehyde	C <sub>9</sub> H <sub>7</sub> NO	145.053	
HILIC(n)	Glaurin	C <sub>16</sub> H <sub>32</sub> O <sub>4</sub>	288.230	
C18(p)	Oxaloacetate	C <sub>4</sub> H <sub>3</sub> O <sub>5</sub>	130.998	
C18(p)	2-Chloroethylphosphonic acid	C <sub>2</sub> H <sub>6</sub> ClO <sub>3</sub> P	143.974	
C18(p)	Dibenzylamine	C <sub>14</sub> H <sub>15</sub> N	197.12	
C18(p)	3-(Aminomethyl)aniline	C <sub>7</sub> H <sub>10</sub> N <sub>2</sub>	122.084	

**Table C.6:** List of compounds discussed in Chapter 5, including the molecular formulae, monoisotopic masses, structures and chromatography/ionization modes of detection. (cont.)

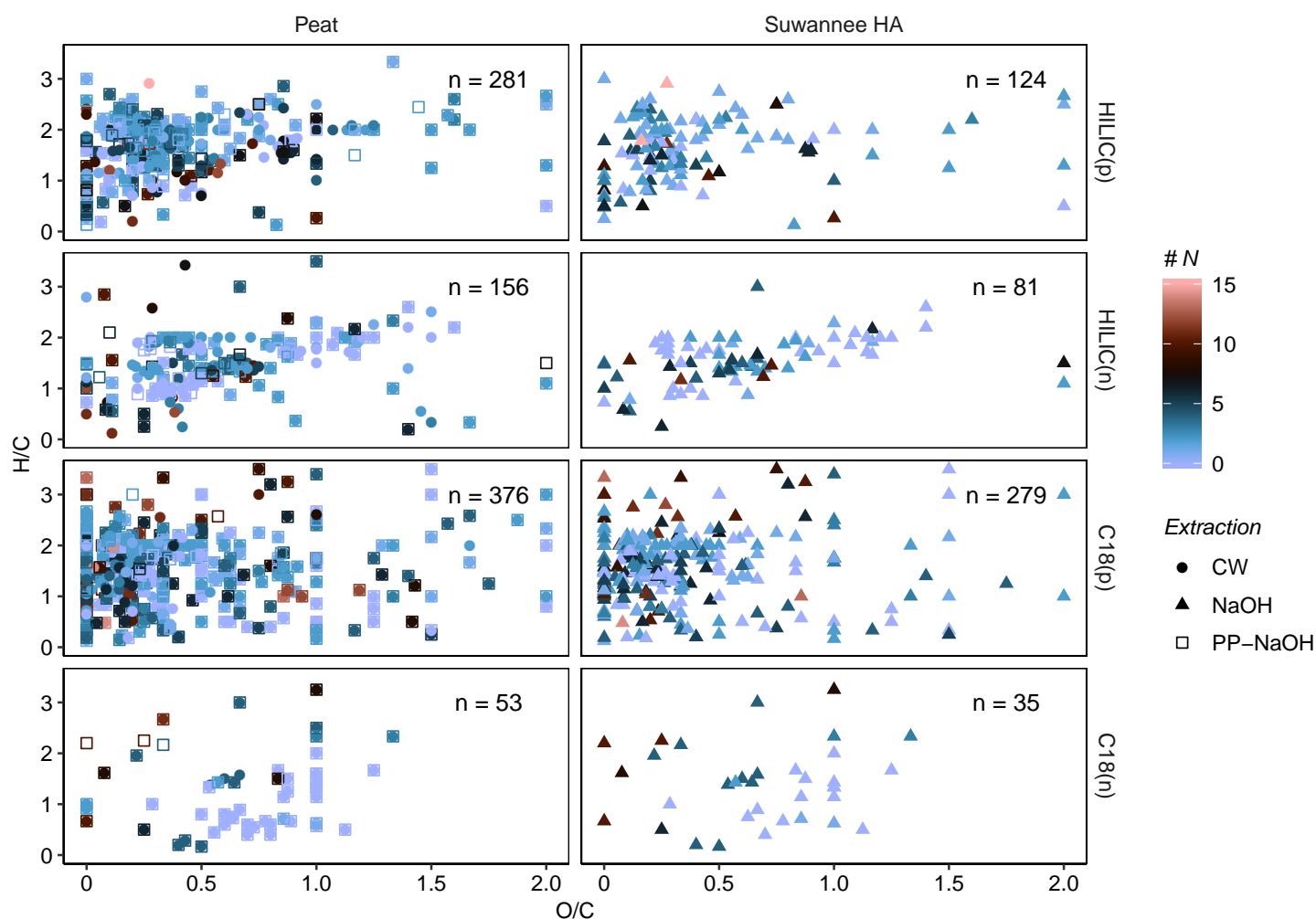
C18(p)	Caprolactam	$C_6H_{11}NO$	113.084	
C18(p)	Aminocaproic acid	$C_6H_{13}NO_2$	131.095	 D1—NH <sub>2</sub>
C18(p)	2-Ethylpiperidine	$C_7H_{15}N$	113.12	
C18(p)	N,N-Diethyl-N'-methylethylene-diamine	$C_7H_{18}N_2$	130.147	
C18(p)	5-Ethyl-2-methylpyridine	$C_8H_{11}N$	121.089	
C18(p)	1,3-Di-tert-butylbenzene	$C_{14}H_{22}$	190.172	
C18(n)	Pyromellitic acid	$C_{10}H_6O_8$	254.006	
C18(n)	(S)-Ureidoglycine	$C_3H_7N_3O_3$	133.049	
C18(p) and (n)	Citramalic acid	$C_5H_8O_5$	148.037	

**Table C.6:** List of compounds discussed in Chapter 5, including the molecular formulae, monoisotopic masses, structures and chromatography/ionization modes of detection. These structures have been manually interpreted by automated searches. Superscripts *a* and *b* denote their presence in pre-incubation, post-resin treated hot water extracts; superscripts *c* and *d* denote their detection in pre-incubation, post-resin treated cold water extracts.

C18(p) <sup>a</sup>	<i>P</i> -[1-[(2-amino-5-pyrimidinyl)methyl]amino]ethyl]-Phosphonic acid	C <sub>7</sub> H <sub>13</sub> N <sub>4</sub> O <sub>3</sub> P	232.179	
C18(p) <sup>b</sup>	4-(dibutylphosphinyl)- $\gamma$ -ethenyl-, ( $\gamma S$ )-Benzenebutanethiol	C <sub>20</sub> H <sub>33</sub> OPS	352.51	
HILIC(p) <sup>c</sup>	<i>N</i> ,3,3-trimethyl- <i>N</i> -(tetrahydro-3-thienyl)-Butanamide	C <sub>11</sub> H <sub>21</sub> NOS	215.134	
C18(p) <sup>d</sup>	10-[(4-methylphenyl)sulfonyl]- 1,4-Dioxo-7, 13-dithia-10-azacyclopentadecane	C <sub>17</sub> H <sub>27</sub> NO <sub>4</sub> S <sub>3</sub>	405.60	

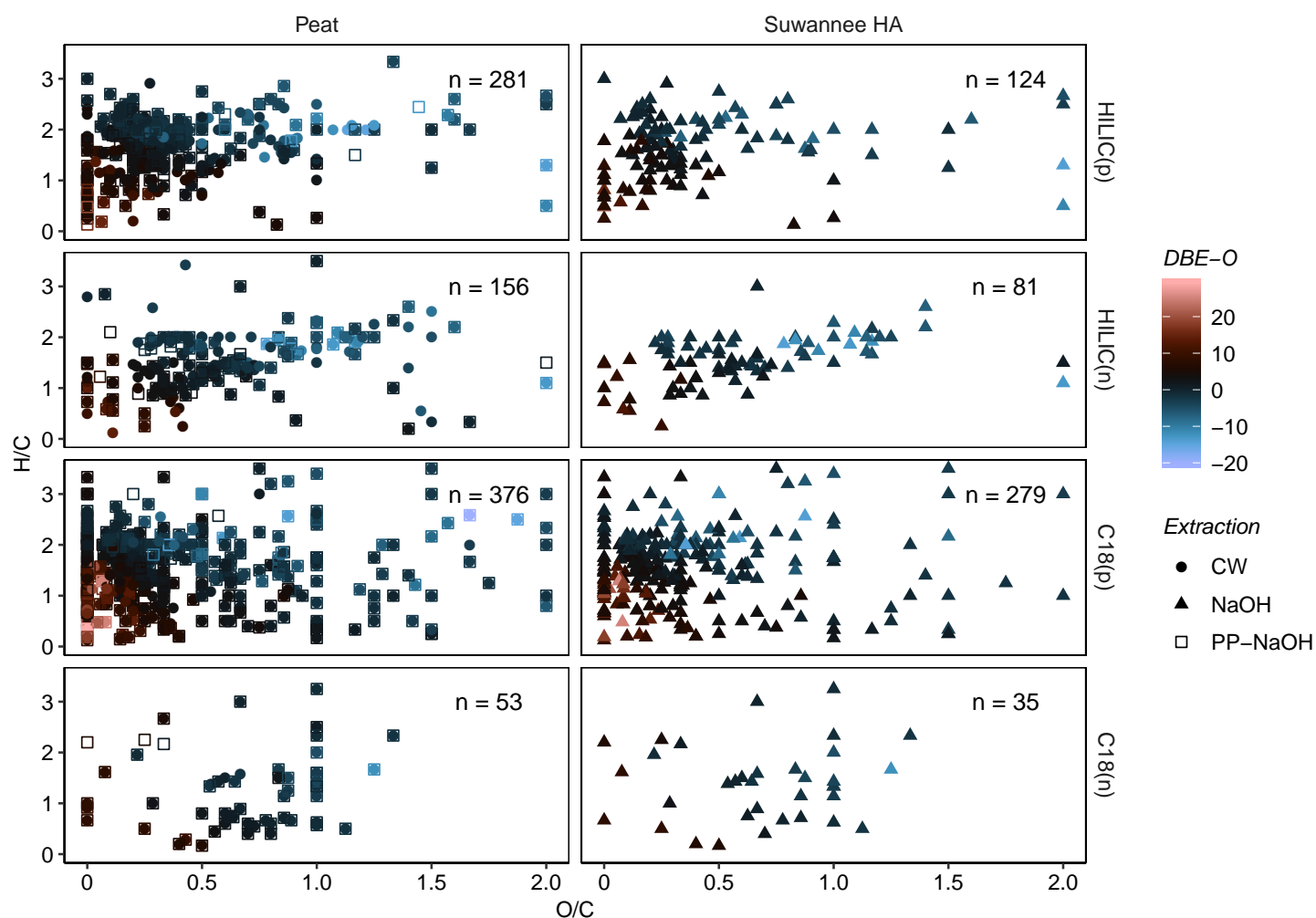


**Figure C.21:** van Krevelen diagrams of the peat and Suwannee River humic acid extracts plotted against the molecular weight (MW) and relative to their chromatography/ionization mode of detection.  $n$  = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes).

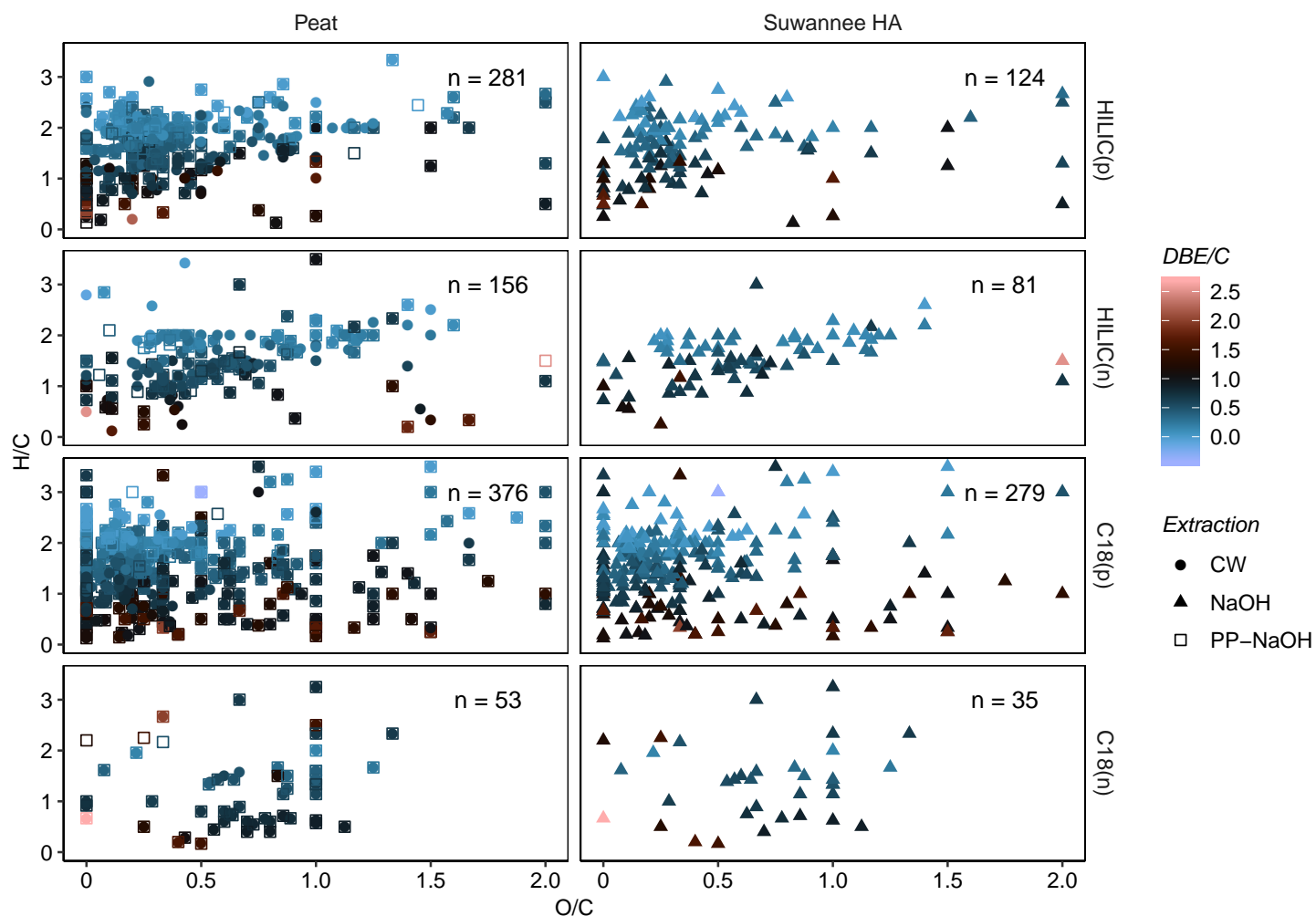


**Figure C.22:** van Krevelen diagrams of the peat and Suwannee River humic acid extracts plotted against the number of nitrogen atoms ( $\#N$ ) and relative to their chromatography/ionization mode of detection.  $n$  = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes).

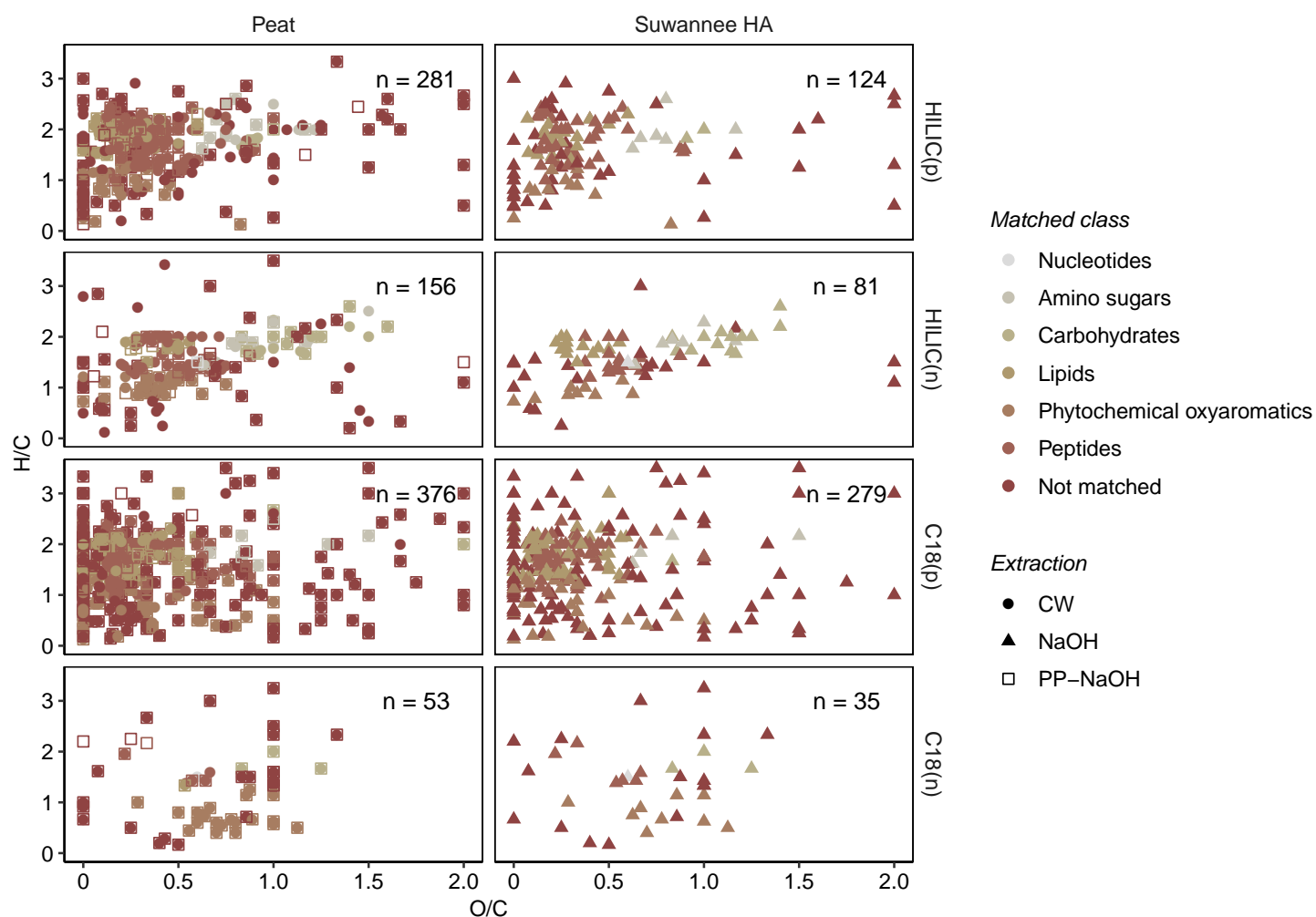




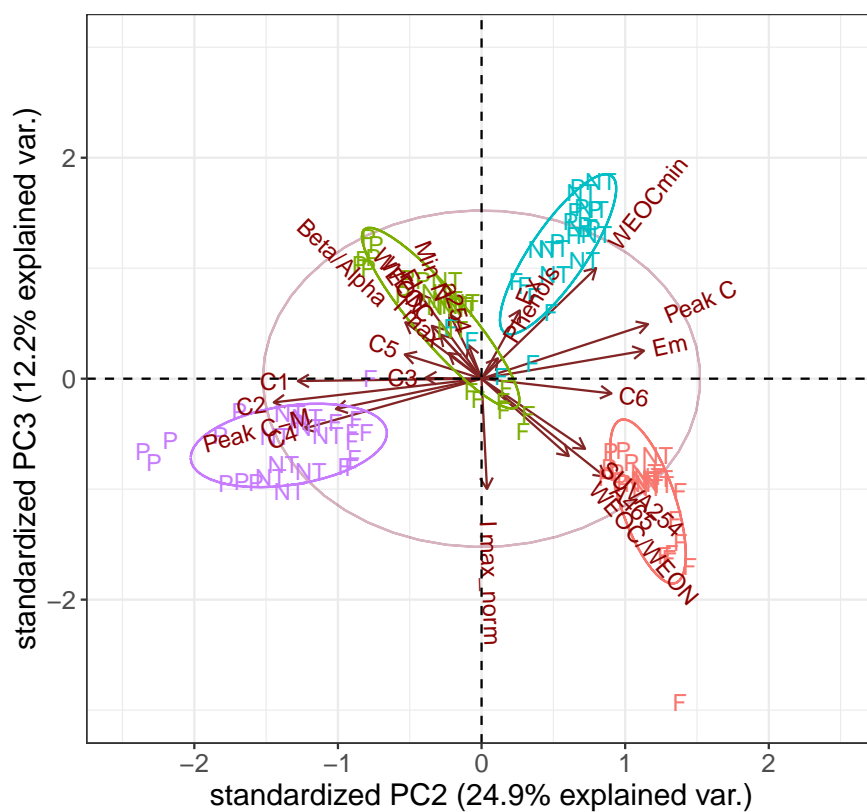
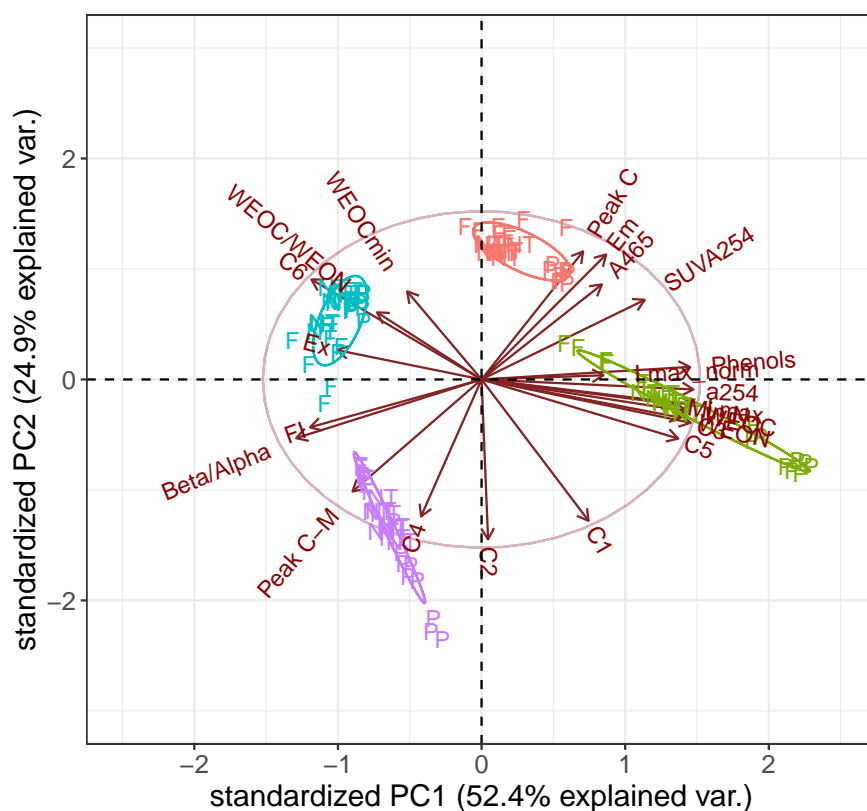
**Figure C.23:** van Krevelen diagrams of the peat and Suwannee River humic acid extracts plotted against DBE-O and relative to their chromatography/ionization mode of detection.  $n$  = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes).



**Figure C.24:** van Krevelen diagrams of the peat and Suwannee River humic acid extracts plotted against DBE/C and relative to their chromatography/ionization mode of detection.  $n$  = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes).



**Figure C.25:** van Krevelen diagrams of compounds classified according to Rivas-Ubach et al. (2018) in the peat and Suwannee River humic acid extracts plotted relative to their chromatography/ionization mode of detection.  $n$  = the number of compounds revealed under the correspondent chromatography/ionization mode (HILIC and C18 in positive (p) and negative (n) ionization modes).



— non-resin treated\_CW   
 — non-resin treated\_HW   
 — post-resin treated\_CW   
 — post-resin treated\_HW

**Figure C.26:** Principal component analysis of all PARAFAC component scores (C1–C6), fluorescence peaks, maximum intensities ( $I_{max}$ ) and relative excitation (Ex) and emission (Em) wavelengths, and all quantitative and qualitative variables investigated in Chapter 4.

Search description: Untargeted Metabolomics workflow: Find and identify the differences between samples.

Performs retention time alignment, unknown compound detection, and compound grouping across all samples. Predicts elemental compositions for all compounds, fills gaps across all samples, and hides chemical background (using Blank samples). Identifies compounds using mzCloud (ddMS2) and ChemSpider (formula or exact mass). Also performs similarity search for all compounds with ddMS2 data using mzCloud. Applies mzLogic algorithm to rank order ChemSpider results. Maps compounds to biological pathways using Metabolika. Applies QC-based batch normalization if QC samples are available. Calculates differential analysis (t-test or ANOVA), determines p-values, adjusted p-values, ratios, fold change, CV, etc.).

-----  
Search ID: HILIC positive  
-----

Created with Discoverer version: 3.0.0.294

[Input Files (6)]

-->Select Spectra (33)

[Select Spectra (33)]

-->Align Retention Times (26)

[Align Retention Times (26)]

-->Detect Compounds (9)

[Detect Compounds (9)]

-->Group Compounds (31)

[Group Compounds (31)]

-->Search mzCloud (22)

-->Assign Compound Annotations (25)

-->Map to BioCyc Pathways (40)

-->Map to KEGG Pathways (41)

-->Fill Gaps (32)

-->Predict Compositions (29)

-->Map to Metabolika Pathways (34)

-->Search ChemSpider (42)

-->Search Mass Lists (36)

-->Search mzVault (37)

[Fill Gaps (32)]  
-->Normalize Areas (30)  
[Map to BioCyc Pathways (40)]  
-->Apply mzLogic (35)  
[Normalize Areas (30)]  
-->Mark Background Compounds (28)  
[Map to Metabolika Pathways (34)]  
-->Apply mzLogic (35)  
[Search ChemSpider (42)]  
-->Apply mzLogic (35)  
[Search Mass Lists (36)]  
-->Apply mzLogic (35)  
[Search mzCloud (22)]  
[Assign Compound Annotations (25)]  
[Apply mzLogic (35)]  
[Map to KEGG Pathways (41)]  
[Mark Background Compounds (28)]  
[Predict Compositions (29)]  
[Search mzVault (37)]  
[Differential Analysis (17)]  
[Descriptive Statistics (38)]

-----  
Processing node 6: Input Files  
-----  
-----

Processing node 33: Select Spectra  
-----

1. General Settings:

- Precursor Selection: Use MS(n - 1) Precursor
- Use Isotope Pattern in Precursor Reevaluation: True
- Provide Profile Spectra: Automatic
- Store Chromatograms: False

## 2. Spectrum Properties Filter:

- Lower RT Limit: 0
- Upper RT Limit: 0
- First Scan: 0
- Last Scan: 0
- Ignore Specified Scans: (not specified)
- Lowest Charge State: 0
- Highest Charge State: 0
- Min. Precursor Mass: 0 Da
- Max. Precursor Mass: 3000 Da
- Total Intensity Threshold: 0
- Minimum Peak Count: 1

## 3. Scan Event Filters:

- Mass Analyzer: Is FTMS
- MS Order: Any
- Activation Type: Is HCD
- Min. Collision Energy: 0
- Max. Collision Energy: 1000
- Scan Type: Any
- Polarity Mode: Is +

## 4. Peak Filters:

- S/N Threshold (FT-only): 1.5

## 5. Replacements for Unrecognized Properties:

- Unrecognized Charge Replacements: 1
- Unrecognized Mass Analyzer Replacements: FTMS
- Unrecognized MS Order Replacements: MS2
- Unrecognized Activation Type Replacements: HCD
- Unrecognized Polarity Replacements: +
- Unrecognized MS Resolution@200 Replacements: 60000
- Unrecognized MSn Resolution@200 Replacements: 30000

-----

Processing node 26: Align Retention Times

-----

1. General Settings:

- Alignment Model: Adaptive curve
  - Alignment Fallback: Use Linear Model
  - Maximum Shift [min]: 2
  - Shift Reference File: True
  - Mass Tolerance: 5 ppm
  - Remove Outlier: True
- 

Processing node 9: Detect Compounds

-----

1. General Settings:

- Mass Tolerance [ppm]: 5 ppm
- Intensity Tolerance [%]: 30
- S/N Threshold: 3
- Min. Peak Intensity: 500000
- Ions:

[2M+ACN+H]+1	[M+Cl]-1
[2M+ACN+Na]+1	[M+FA-H]-1
[2M+FA-H]-1	[M+H]+1
[2M+H]+1	[M+H+K]+2
[2M+K]+1	[M+H+MeOH]+1
[2M+Na]+1	[M+H+Na]+2
[2M+NH4]+1	[M+H+NH4]+2
[2M-H]-1	[M+H-H2O]+1
[2M-H+HAc]-1	[M+H-NH3]+1
[M+2H]+2	[M+K]+1
[M+ACN+2H]+2	[M+Na]+1
[M+ACN+H]+1	[M+NH4]+1
[M+ACN+Na]+1	[M-2H]-2



[M-2H+K]-1	[M-H+HAc]-1
[M-H]-1	[M-H-H <sub>2</sub> O]-1
- Base Ions (Advanced):	
[2M+H]+1	[M+H+Na]+2
[2M+K]+1	[M+H+NH <sub>4</sub> ]+2
[2M+Na]+1	[M+H-H <sub>2</sub> O]+1
[2M+NH <sub>4</sub> ]+1	[M+H-NH <sub>3</sub> ]+1
[M+2H]+2	[M+K]+1
[M+3H]+3	[M+Na]+1
[M+H]+1	[M+NH <sub>4</sub> ]+1
[M+H+K]+2	

- Min. Element Counts: C H

- Max. Element Counts: C90 H100 N15 O20 P S5

## 2. Peak Detection:

- Filter Peaks: True

- Max. Peak Width [min]: 0.5

- Remove Singlets: True

- Min. # Scans per Peak: 5

- Min. # Isotopes: 1

-----  
Processing node 31: Group Compounds  
-----

## 1. Compound Consolidation:

- Mass Tolerance: 5 ppm

- RT Tolerance [min]: 0.2

## 2. Fragment Data Selection:

- Preferred Ions:

[2M+H]+1	[M+3H]+3
[2M+Na]+1	[M+H]+1
[2M+NH <sub>4</sub> ]+1	[M+H+Na]+2
[M+2H]+2	[M+H+NH <sub>4</sub> ]+2

[M+H-H<sub>2</sub>O]+1

[M+Na]+1

[M+H-NH<sub>3</sub>]+1

[M+NH<sub>4</sub>]+

-----  
Processing node 22: Search mzCloud  
-----

1. Search Settings:

- Compound Classes: All
  - Match Ion Activation Type: True
  - Match Ion Activation Energy: Match with Tolerance
  - Ion Activation Energy Tolerance: 20
  - Apply Intensity Threshold: True
  - Precursor Mass Tolerance: 10 ppm
  - FT Fragment Mass Tolerance: 10 ppm
  - IT Fragment Mass Tolerance: 0.4 Da
  - Identity Search: HighChem HighRes
  - Similarity Search: Similarity Forward
  - Library: Reference
  - Post Processing: Recalibrated
  - Match Factor Threshold: 50
  - Max. # Results: 10
- 

Processing node 25: Assign Compound Annotations  
-----

1. General Settings:

- Mass Tolerance: 5 ppm

2. Data Sources:

- Data Source #1: Predicted Compositions
- Data Source #2: MassList Search
- Data Source #3: Metabolika Search
- Data Source #4: BioCyc Search
- Data Source #5: mzVault Search

---

Processing node 40: Map to BioCyc Pathways

---

1. Search Settings:

- BioCyc Database/organism to be searched: MetaCyc (META)
- Search Mode: By Formula or Mass

2. By Mass Search Settings:

- Mass Tolerance: 5 ppm

3. By Formula Search Settings:

- Max. # of Predicted Compositions to be searched per Compound: 3

4. Display Settings:

- Max. # Pathways in 'Pathways' column: 20
- 

Processing node 35: Apply mzLogic

---

1. Search Settings:

- FT Fragment Mass Tolerance: 10 ppm
  - IT Fragment Mass Tolerance: 0.4 Da
  - Max. # Compounds: 0
  - Max. # mzCloud Similarity Results to consider per Compound: 10
  - Match Factor Threshold: 30
- 

Processing node 41: Map to KEGG Pathways

---

1. Search Settings:

- Search Mode: By Formula or Mass

2. By Mass Search Settings:

- Mass Tolerance: 5 ppm

3. By Formula Search Settings:

- Max. # of Predicted Compositions to be searched per Compound: 3

#### 4. Display Settings:

- Max. # Pathways in 'Pathways' column: 20

---

#### Processing node 32: Fill Gaps

---

#### 1. General Settings:

- Mass Tolerance: 5 ppm
- S/N Threshold: 1.5
- Use Real Peak Detection: True

---

#### Processing node 30: Normalize Areas

---

#### 1. QC-based Area Correction:

- Regression Model: Cubic Spline
- Min. QC Coverage [%]: 50
- Max. QC Area RSD [%]: 30
- Max. # Files Between QC Files: 20

#### 2. Area Normalization:

- Normalization Type: None
- Exclude Blanks: True

---

#### Processing node 28: Mark Background Compounds

---

#### 1. General Settings:

- Max. Sample/Blank: 5
- Max. Blank/Sample: 0
- Hide Background: True

---

#### Processing node 29: Predict Compositions

---

#### 1. Prediction Settings:

- Mass Tolerance: 5 ppm

- Min. Element Counts: C H
- Max. Element Counts: C90 H100 N15 O20 P S5
- Min. RDBE: 0
- Max. RDBE: 40
- Min. H/C: 0.1
- Max. H/C: 4
- Max. # Candidates: 10
- Max. # Internal Candidates: 200

#### 2. Pattern Matching:

- Intensity Tolerance [%]: 30
- Intensity Threshold [%]: 0.1
- S/N Threshold: 3
- Min. Spectral Fit [%]: 30
- Min. Pattern Cov. [%]: 90
- Use Dynamic Recalibration: True

#### 3. Fragments Matching:

- Use Fragments Matching: True
- Mass Tolerance: 5 ppm
- S/N Threshold: 3

-----  
Processing node 34: Map to Metabolika Pathways  
-----

#### 1. Search Settings:

- Search Mode: By Formula or Mass

#### 2. By Mass Search Settings:

- Mass Tolerance: 5 ppm

#### 3. By Formula Search Settings:

- Max. # of Predicted Compositions to be searched per Compound: 3

#### 4. Display Settings:

- Max. # Pathways in 'Pathways' column: 20

-----  
Processing node 42: Search ChemSpider  
-----

1. Search Settings:

- Database(s):

ACToR: Aggregated Computational Toxicology Resource

BioCyc

ChEBI

E. coli Metabolome Database

EAWAG Biocatalysis/Biodegradation Database

EPA DSSTox

FDA UNII - NLM

KEGG

LipidMAPS

MCISB

PlantCyc

SMPDB Small Molecule Pathway Database

Yeast Metabolome Database

- Search Mode: By Formula or Mass

- Mass Tolerance: 5 ppm

- Max. # of results per compound: 100

- Max. # of Predicted Compositions to be searched per Compound: 3

- Result Order (for Max. # of results per compound): Order By Reference Count (DESC)

2. Predicted Composition Annotation:

- Check All Predicted Compositions: False  
-----

Processing node 36: Search Mass Lists  
-----

1. Search Settings:

- Mass Lists: Extractables and Leachables HRAM Compound Database.massList |  
TV.massList | OpenVK.massList

- Mass Tolerance: 5 ppm

- Use Retention Time: True

- RT Tolerance [min]: 2

-----  
Processing node 37: Search mzVault

-----  
1. Search Settings:

- mzVault Library: Bamba lab 34 lipid mediators library stepped NCE 10 30 45.db | Bamba lab 598 polar metabolites stepped NCE 10 30 45.db | mzVault May 2018.db | Lipids-Aug18.db | MoNA-export-LC-MS-Jan19.db | J0022-ref-H-T001-(1).db

- Max. # Results: 10

- Match Factor Threshold: 50

- Search Algorithm: HighChem HighRes

- Match Analyzer Type: False

- IT Fragment Mass Tolerance: 0.4 Da

- FT Fragment Mass Tolerance: 10 ppm

- Use Retention Time: False

- Precursor Mass Tolerance: 10 ppm

- Apply Intensity Threshold: True

- Match Ionization Method: False

- Ion Activation Energy Tolerance: 20

- Match Ion Activation Energy: Any

- Match Ion Activation Type: False

- Compound Classes: All

- Remove Precursor Ion: True

- RT Tolerance [min]: 2

-----  
Processing node 17: Differential Analysis

-----  
1. General Settings:

- Log10 Transform Values: True

-----  
Processing node 38: Descriptive Statistics

-----  
Search ID: HILIC negative, C18 positive, C18 negative  
-----

Created with Discoverer version: 3.0.0.294

[Input Files (6)]

-->Select Spectra (33)

[Select Spectra (33)]

-->Align Retention Times (26)

[Align Retention Times (26)]

-->Detect Compounds (9)

[Detect Compounds (9)]

-->Group Compounds (31)

[Group Compounds (31)]

-->Search mzCloud (22)

-->Assign Compound Annotations (25)

-->Map to BioCyc Pathways (40)

-->Map to KEGG Pathways (41)

-->Fill Gaps (32)

-->Predict Compositions (29)

-->Map to Metabolika Pathways (34)

-->Search Mass Lists (36)

-->Search mzVault (37)

[Fill Gaps (32)]

-->Normalize Areas (30)

[Map to BioCyc Pathways (40)]

-->Apply mzLogic (35)

[Normalize Areas (30)]

-->Mark Background Compounds (28)

[Map to Metabolika Pathways (34)]



-->Apply mzLogic (35)  
[Search Mass Lists (36)]  
-->Apply mzLogic (35)  
[Search mzCloud (22)]  
[Assign Compound Annotations (25)]  
[Apply mzLogic (35)]  
[Map to KEGG Pathways (41)]  
[Mark Background Compounds (28)]  
[Predict Compositions (29)]  
[Search mzVault (37)]  
[Differential Analysis (17)]  
[Descriptive Statistics (38)]

-----  
Processing node 6: Input Files  
-----  
-----

Processing node 33: Select Spectra  
-----

1. General Settings:

- Precursor Selection: Use MS(n - 1) Precursor
- Use Isotope Pattern in Precursor Reevaluation: True
- Provide Profile Spectra: Automatic
- Store Chromatograms: False

2. Spectrum Properties Filter:

- Lower RT Limit: 0
- Upper RT Limit: 0
- First Scan: 0
- Last Scan: 0
- Ignore Specified Scans: (not specified)
- Lowest Charge State: 0

- Highest Charge State: 0
- Min. Precursor Mass: 0 Da
- Max. Precursor Mass: 1000 Da
- Total Intensity Threshold: 0
- Minimum Peak Count: 1

### 3. Scan Event Filters:

- Mass Analyzer: Is FTMS
- MS Order: Any
- Activation Type: Is HCD
- Min. Collision Energy: 0
- Max. Collision Energy: 1000
- Scan Type: Any
- Polarity Mode: (not specified)

### 4. Peak Filters:

- S/N Threshold (FT-only): 1.5

### 5. Replacements for Unrecognized Properties:

- Unrecognized Charge Replacements: 1
- Unrecognized Mass Analyzer Replacements: FTMS
- Unrecognized MS Order Replacements: MS2
- Unrecognized Activation Type Replacements: HCD
- Unrecognized Polarity Replacements: +
- Unrecognized MS Resolution@200 Replacements: 60000
- Unrecognized MSn Resolution@200 Replacements: 30000

-----  
Processing node 26: Align Retention Times  
-----

### 1. General Settings:

- Alignment Model: Adaptive curve
- Alignment Fallback: Use Linear Model
- Maximum Shift [min]: 2

- Shift Reference File: True
- Mass Tolerance: 5 ppm
- Remove Outlier: True

-----

Processing node 9: Detect Compounds

-----

1. General Settings:

- Mass Tolerance [ppm]: 5 ppm
- Intensity Tolerance [%]: 30
- S/N Threshold: 3
- Min. Peak Intensity: 500000
- Ions:

[2M+ACN+H]+1	[M+H]+1
[2M+ACN+Na]+1	[M+H+K]+2
[2M+FA-H]-1	[M+H+MeOH]+1
[2M+H]+1	[M+H+Na]+2
[2M+K]+1	[M+H+NH4]+2
[2M+Na]+1	[M+H-H2O]+1
[2M+NH4]+1	[M+H-NH3]+1
[2M-H]-1	[M+K]+1
[2M-H+HAc]-1	[M+Na]+1
[M+2H]+2	[M+NH4]+1
[M+ACN+2H]+2	[M-2H]-2
[M+ACN+H]+1	[M-2H+K]-1
[M+ACN+Na]+1	[M-H]-1
[M+Cl]-1	[M-H+HAc]-1
[M+FA-H]-1	[M-H-H2O]-1

- Base Ions: [M+H]+1; [M-H]-1
- Min. Element Counts: C H
- Max. Element Counts: C90 H100 N15 O20 P S5

## 2. Peak Detection:

- Filter Peaks: True
- Max. Peak Width [min]: 0.5
- Remove Singlets: True
- Min. # Scans per Peak: 5
- Min. # Isotopes: 1

-----  
Processing node 31: Group Compounds  
-----

## 1. Compound Consolidation:

- Mass Tolerance: 5 ppm
- RT Tolerance [min]: 0.2

## 2. Fragment Data Selection:

- Preferred Ions: [M+FA-H]-1; [M+H]+1; [M-H]-1; [M-H-H2O]-1

-----  
Processing node 22: Search mzCloud  
-----

## 1. Search Settings:

- Compound Classes: All
- Match Ion Activation Type: True
- Match Ion Activation Energy: Match with Tolerance
- Ion Activation Energy Tolerance: 20
- Apply Intensity Threshold: True
- Precursor Mass Tolerance: 10 ppm
- FT Fragment Mass Tolerance: 10 ppm
- IT Fragment Mass Tolerance: 0.4 Da
- Identity Search: HighChem HighRes
- Similarity Search: Similarity Forward
- Library: Reference
- Post Processing: Recalibrated

- Match Factor Threshold: 50

- Max. # Results: 10

-----  
Processing node 25: Assign Compound Annotations  
-----

1. General Settings:

- Mass Tolerance: 5 ppm

2. Data Sources:

- Data Source #1: Predicted Compositions

- Data Source #2: MassList Search

- Data Source #3: Metabolika Search

- Data Source #4: BioCyc Search

- Data Source #5: mzVault Search

-----  
Processing node 40: Map to BioCyc Pathways  
-----

1. Search Settings:

- BioCyc Database/organism to be searched: MetaCyc (META)

- Search Mode: By Formula or Mass

2. By Mass Search Settings:

- Mass Tolerance: 5 ppm

3. By Formula Search Settings:

- Max. # of Predicted Compositions to be searched per Compound: 3

4. Display Settings:

- Max. # Pathways in 'Pathways' column: 20

-----  
Processing node 35: Apply mzLogic  
-----

1. Search Settings:

- FT Fragment Mass Tolerance: 10 ppm

- IT Fragment Mass Tolerance: 0.4 Da
  - Max. # Compounds: 0
  - Max. # mzCloud Similarity Results to consider per Compound: 10
  - Match Factor Threshold: 30
- 

#### Processing node 41: Map to KEGG Pathways

-----

##### 1. Search Settings:

- Search Mode: By Formula or Mass

##### 2. By Mass Search Settings:

- Mass Tolerance: 5 ppm

##### 3. By Formula Search Settings:

- Max. # of Predicted Compositions to be searched per Compound: 3

##### 4. Display Settings:

- Max. # Pathways in 'Pathways' column: 20
- 

#### Processing node 32: Fill Gaps

-----

##### 1. General Settings:

- Mass Tolerance: 5 ppm
  - S/N Threshold: 1.5
  - Use Real Peak Detection: True
- 

#### Processing node 30: Normalize Areas

-----

##### 1. QC-based Area Correction:

- Regression Model: Cubic Spline
- Min. QC Coverage [%]: 50
- Max. QC Area RSD [%]: 30
- Max. # Files Between QC Files: 20

## 2. Area Normalization:

- Normalization Type: None
- Exclude Blanks: True

-----  
Processing node 28: Mark Background Compounds  
-----

## 1. General Settings:

- Max. Sample/Blank: 5
- Max. Blank/Sample: 0
- Hide Background: True

-----  
Processing node 29: Predict Compositions  
-----

## 1. Prediction Settings:

- Mass Tolerance: 5 ppm
- Min. Element Counts: C H
- Max. Element Counts: C90 H100 N15 O20 P S5
- Min. RDBE: 0
- Max. RDBE: 40
- Min. H/C: 0.1
- Max. H/C: 4
- Max. # Candidates: 10
- Max. # Internal Candidates: 200

## 2. Pattern Matching:

- Intensity Tolerance [%]: 30
- Intensity Threshold [%]: 0.1
- S/N Threshold: 3
- Min. Spectral Fit [%]: 30
- Min. Pattern Cov. [%]: 90
- Use Dynamic Recalibration: True

### 3. Fragments Matching:

- Use Fragments Matching: True
  - Mass Tolerance: 5 ppm
  - S/N Threshold: 3
- 

### Processing node 34: Map to Metabolika Pathways

-----

#### 1. Search Settings:

- Search Mode: By Formula or Mass

#### 2. By Mass Search Settings:

- Mass Tolerance: 5 ppm

#### 3. By Formula Search Settings:

- Max. # of Predicted Compositions to be searched per Compound: 3

#### 4. Display Settings:

- Max. # Pathways in 'Pathways' column: 20
- 

### Processing node 36: Search Mass Lists

-----

#### 1. Search Settings:

- Mass Lists: Extractables and Leachables HRAM Compound Database.massList | OpenVK.massList | TV.massList
  - Mass Tolerance: 5 ppm
  - Use Retention Time: True
  - RT Tolerance [min]: 2
- 

### Processing node 37: Search mzVault

-----

#### 1. Search Settings:

- mzVault Library Bamba lab 34 lipid mediators library stepped NCE 10 30 45.db | Bamba lab 598 polar metabolites stepped NCE 10 30 45.db | mzVault May 2018.db | J0022-ref-H-T001-(1).db



- Max. # Results: 10
- Match Factor Threshold: 50
- Search Algorithm: HighChem HighRes
- Match Analyzer Type: False
- IT Fragment Mass Tolerance: 0.4 Da
- FT Fragment Mass Tolerance: 10 ppm
- Use Retention Time: False
- Precursor Mass Tolerance: 10 ppm
- Apply Intensity Threshold: True
- Match Ionization Method: False
- Ion Activation Energy Tolerance: 20
- Match Ion Activation Energy: Any
- Match Ion Activation Type: False
- Compound Classes: All
- Remove Precursor Ion: True
- RT Tolerance [min]: 2

-----  
Processing node 17: Differential Analysis  
-----

1. General Settings:

- Log10 Transform Values: True

-----  
Processing node 38: Descriptive Statistics  
-----

This page is intentionally left blank

# Bibliography

- Aceto, M., Sarzanini, C., Abollino, O., and Mentasti, E. (1995). Ion chromatographic separation of alkylsulphonic acids with conductivity detection. *Chromatographia*, 41(5):445–449.
- Adu, J. K. and Oades, J. M. (1978). Physical factors influencing decomposition of organic materials in soil aggregates. *Soil Biology and Biochemistry*, 10(2):109–115.
- Afcharian, A., Levi, Y., Kiene, L., and Scribe, P. (1997). Fractionation of dissolved organic matter from surface waters using macroporous resins. *Water Research*, 31(12):2989–2996.
- Ågren, G. I. (2000). Temperature dependence of old soil organic matter. *AMBIO: A Journal of the Human Environment*, 29(1):55.
- Ågren, G. I. and Bosatta, E. (1998). *Theoretical Ecosystem Ecology: Understanding Element Cycles*. Cambridge University Press, Cambridge, UK.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.
- Albani, J. R. (2008). *Principles and Applications of Fluorescence Spectroscopy*. John Wiley & Sons.
- Almendros, G. and Dorado, J. (1999). Molecular characteristics related to the biodegradability of humic acid preparations. *European Journal of Soil Science*, 50(2):227–236.
- Alva, A. K., Sumner, M. E., and Miller, W. P. (1991). Relationship between ionic strength and electrical conductivity for soil solutions. *Soil Science*, 152(4).
- Amelung, W. (2003). Nitrogen biomarkers and their fate in soil. *Journal of Plant Nutrition and Soil Science*, 166(6):677–686.

- Andersen, C. M. and Bro, R. (2003). Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics*, 17(4):200–215.
- Angst, G., Mueller, K. E., Kögel-Knabner, I., Freeman, K. H., and Mueller, C. W. (2017). Aggregation controls the stability of lignin and lipids in clay-sized particulate and mineral associated organic matter. *Biogeochemistry*, 132(3):307–324.
- Antizar-Ladislao, B., Lopez-Real, J., and Beck, A. J. (2006). Investigation of organic matter dynamics during in-vessel composting of an aged coal-tar contaminated soil using fluorescence excitation-emission spectroscopy. *Chemosphere*, 64(5):839–847.
- Apostel, C., Dippold, M., and Kuzyakov, Y. (2015). Biochemistry of hexose and pentose transformations in soil analyzed by position-specific labeling and  $^{13}\text{C}$ -PLFA. *Soil Biology and Biochemistry*, 80:199–208.
- Atkeson, C. G., Moore, A. W., and Schaal, S. (1997). Locally Weighted Learning. *Artificial Intelligence Review*, 11(1):11–73.
- Auger, G., van Heijenoort, J., Vederas, J. C., and Blanot, D. (1996). Effect of analogues of diaminopimelic acid on the meso-diaminopimelate-adding enzyme from *Escherichia coli*. *FEBS Letters*, 391(1-2):171–174.
- Avneri-Katz, S., Young, R. B., McKenna, A. M., Chen, H., Corilo, Y. E., Polubesova, T., Borch, T., and Chofetz, B. (2017). Adsorptive fractionation of dissolved organic matter (DOM) by mineral soil: Macroscale approach and molecular insight. *Organic Geochemistry*, 103:113–124.
- Badalucco, L. and Nannipieri, P. (2007). Nutrient transformations in the rhizosphere. In Pinton, R., Varanini, Z., and Nannipieri, P., editors, *The rhizosphere biochemistry and organic substances at the soil-plant interface*, pages 111–133. CRC Press, Boca Raton, FL, USA.
- Bailey, V. L., Smith, A. P., Tfaily, M., Fansler, S. J., and Bond-Lamberty, B. (2017). Differences in soluble organic carbon chemistry in pore waters sampled from different pore size domains. *Soil Biology and Biochemistry*, 107:133–143.
- Baker, A., Tipping, E., Thacker, S. A., and Gondar, D. (2008). Relating dissolved organic matter fluorescence and functional properties. *Chemosphere*, 73(11):1765–1772.
- Baker, J. M., Ochsner, T. E., Venterea, R. T., and Griffis, T. J. (2007). Tillage and soil carbon sequestration—What do we really know? *Agriculture, Ecosystems & Environment*, 118(1):1–5.

- Baldock, J. A. and Broos, K. (2012). Soil organic matter. In Huang, P. M., Li, Y., and Sumner, M. E., editors, *Handbook of Soil Sciences: Properties and Processes*, pages 11–52. CRC Press, Boca Raton, FL, USA, 2<sup>nd</sup> edition.
- Baldock, J. A., Oades, J. M., Waters, A. G., Peng, X., Vassallo, A. M., and Wilson, M. A. (1992). Aspects of the chemical structure of soil organic materials as revealed by solid-state  $^{13}\text{C}$  NMR spectroscopy. *Biogeochemistry*, 16(1):1–42.
- Bandowe, B. A. M., Leimer, S., Meusel, H., Velescu, A., Dassen, S., Eisenhauer, N., Hoffmann, T., Oelmann, Y., and Wilcke, W. (2019). Plant diversity enhances the natural attenuation of polycyclic aromatic compounds (PAHs and oxygenated PAHs) in grassland soils. *Soil Biology and Biochemistry*, 129:60–70.
- Baran, R., Bowen, B. P., Price, M. N., Arkin, A. P., Deutschbauer, A. M., and Northen, T. R. (2013). Metabolic Footprinting of Mutant Libraries to Map Metabolite Utilization to Genotype. *ACS Chemical Biology*, 8(1):189–199.
- Baran, R., Brodie, E. L., Mayberry-Lewis, J., Hummel, E., Da Rocha, U. N., Chakraborty, R., Bowen, B. P., Karaoz, U., Cadillo-Quiroz, H., Garcia-Pichel, F., and Northen, T. R. (2015). Exometabolite niche partitioning among sympatric soil bacteria. *Nature Communications*, 6:8289.
- Barré, P., Eglin, T., Christensen, B. T., Ciais, P., Houot, S., Kätterer, T., Van Oort, F., Peylin, P., Poulton, P. R., and Romanenkov, V. (2010). Quantifying and isolating stable soil organic carbon using long-term bare fallow experiments. *Biogeosciences*, 2010(7):3839–3850.
- Bashir, M. J. K., Aziz, H. A., Yusoff, M. S., Aziz, S. Q., and Mohajeri, S. (2010). Stabilized sanitary landfill leachate treatment using anionic resin: Treatment optimization by response surface methodology. *Journal of Hazardous Materials*, 182(1):115–122.
- Bayer, C., Martin-Neto, L., Mielniczuk, J., and Ceretta, C. A. (2000). Effect of no-till cropping systems on soil organic matter in a sandy clay loam Acrisol from Southern Brazil monitored by electron spin resonance and nuclear magnetic resonance. *Soil and Tillage Research*, 53(2):95–104.
- Bayer, C., Martin-Neto, L., Mielniczuk, J., Pavinato, A., and Dieckow, J. (2006). Carbon sequestration in two Brazilian Cerrado soils under no-till. *Soil and Tillage Research*, 86(2):237–245.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

- Biedermannova, L., Riley, K. E., Berka, K., Hobza, P., and Vondrasek, J. (2008). Another role of proline: stabilization interactions in proteins and protein complexes concerning proline and tryptophane. *Physical Chemistry Chemical Physics*, 10(42):6350–6359.
- Birch, H. F. (1958). The effect of soil drying on humus decomposition and nitrogen availability. *Plant and Soil*, 10(1):9–31.
- Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin Biosynthesis. *Annual Review of Plant Biology*, 54(1):519–546.
- Bolan, N. S., Adriano, D. C., Kunhikrishnan, A., James, T., McDowell, R., and Senesi, N. (2011). Dissolved Organic Matter: Biogeochemistry, Dynamics, and Environmental Significance in Soils. In Sparks, D. L., editor, *Advances in agronomy*, volume 110, chapter 1, pages 1–75. Academic Press.
- Bolto, B., Dixon, D., Eldridge, R., King, S., and Linge, K. (2002). Removal of natural organic matter by ion exchange. *Water Research*, 36(20):5057–5065.
- Boone, C. and Adamec, J. (2016). 10 – Top-Down Proteomics. In Ciborowski, P. and Silberring, J., editors, *Proteomic Profiling and Analytical Chemistry*, pages 175–191. Elsevier, Boston, 2<sup>nd</sup> edition.
- Borken, W. and Matzner, E. (2009). Reappraisal of drying and wetting effects on C and N mineralization and fluxes in soils. *Global Change Biology*, 15(4):808–824.
- Bottner, P. (1985). Response of microbial biomass to alternate moist and dry conditions in a soil incubated with <sup>14</sup>C- and <sup>15</sup>N-labelled plant material. *Soil Biology and Biochemistry*, 17(3):329–337.
- Bouckaert, R. R., Frank, E., Hall, M. A., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2010). WEKA—Experiences with a Java Open-Source Project. *The Journal of Machine Learning Research*, 11:2533–2541.
- Boyer, J. N. and Groffman, P. M. (1996). Bioavailability of water extractable organic carbon fractions in forest and agricultural soil profiles. *Soil Biology and Biochemistry*, 28(6):783–790.
- Boyer, T. H. and Singer, P. C. (2008). Stoichiometry of Removal of Natural Organic Matter by Ion Exchange. *Environmental Science & Technology*, 42(2):608–613.
- Boyer, T. H., Singer, P. C., and Aiken, G. R. (2008). Removal of Dissolved Organic Matter by Anion Exchange: Effect of Dissolved Organic Matter Properties. *Environmental Science & Technology*, 42(19):7431–7437.

- Bracewell, J. M. and Robertson, G. W. (1984). Quantitative comparison of the nitrogen-containing pyrolysis products and amino acid composition of soil humic acids. *Journal of Analytical and Applied Pyrolysis*, 6(1):19–29.
- Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24(2):123–140.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.
- Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2):149–171.
- Brockman, S. A., Roden, E. V., and Hegeman, A. D. (2018). Van Krevelen diagram visualization of high resolution-mass spectrometry metabolomics data with OpenVanKrevelen. *Metabolomics*, 14(4):48.
- Bronick, C. J. and Lal, R. (2005). Soil structure and management: a review. *Geoderma*, 124(1):3–22.
- Bruderer, R., Bernhardt, O. M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D., and Reiter, L. (2017). Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Molecular & Cellular Proteomics*, 16(12):2296–2309.
- Bu, X., Wang, L., Ma, W., Yu, X., McDowell, W. H., and Ruan, H. (2010). Spectroscopic characterization of hot-water extractable organic matter from soils under four different vegetation types along an elevation gradient in the Wuyi Mountains. *Geoderma*, 159(1):139–146.
- Buckel, W. and Barker, H. A. (1974). Two Pathways of Glutamate Fermentation by Anaerobic Bacteria. *Journal of Bacteriology*, 117(3):1248–1260.
- Burdick, D. S. (1995). An introduction to tensor products with applications to multiway data analysis. *Chemometrics and Intelligent Laboratory Systems*, 28(2):229–237.
- Burford, J. R. and Bremner, J. M. (1975). Relationships between the denitrification capacities of soils and total, water-soluble and readily decomposable soil organic matter. *Soil Biology and Biochemistry*, 7(6):389–394.
- Buszewski, B. and Noga, S. (2012). Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Analytical and Bioanalytical Chemistry*, 402(1):231–247.

- Cabrera, M. L. and Beare, M. H. (1993). Alkaline persulfate oxidation for determining total nitrogen in microbial biomass extracts. *Soil Science Society of America Journal*, 57(4):1007–1012.
- Caetano, M., Valderrama, C., Farran, A., and Cortina, J. L. (2009). Phenol removal from aqueous solution by adsorption and ion exchange mechanisms onto polymeric resins. *Journal of Colloid and Interface Science*, 338(2):402–409.
- Cambardella, C. A. and Elliott, E. T. (1992). Particulate Soil Organic-Matter Changes across a Grassland Cultivation Sequence. *Soil Science Society of America Journal*, 56:777–783.
- Cambardella, C. A. and Elliott, E. T. (1993). Carbon and Nitrogen Distribution in Aggregates from Cultivated and Native Grassland Soils. *Soil Science Society of America Journal*, 57:1071–1076.
- Cao, X. and Schmidt-Rohr, K. (2018). Abundant nonprotonated aromatic and oxygen-bonded carbons make humic substances distinct from biopolymers. *Environmental Science & Technology Letters*, 5:476–480.
- Carmona, M., Lucas, A. D., Valverde, J. L., Velasco, B., and Rodríguez, J. F. (2006). Combined adsorption and ion exchange equilibrium of phenol on Amberlite IRA-420. *Chemical Engineering Journal*, 117(2):155–160.
- Caron, J., Espindola, C. R., and Angers, D. A. (1996). Soil Structural Stability during Rapid Wetting: Influence of Land Use on Some Aggregate Properties. *Soil Science Society of America Journal*, 60:901–908.
- Carter, M. R. and Gregorich, E. G. (2008). *Soil sampling and methods of analysis*. Canadian Society of Soil Science Publication, CRC Press, Boca Raton, FL, USA, 2<sup>nd</sup> edition.
- Casado, P. and Cutillas, P. R. (2011). A Self-validating Quantitative Mass Spectrometry Method for Assessing the Accuracy of High-content Phosphoproteomic Experiments. *Molecular & Cellular Proteomics*, 10(1):M110.003079, 1–11.
- Chantigny, M. H. (2003). Dissolved and water-extractable organic matter in soils: A review on the influence of land use and management practices. *Geoderma*, 113(3-4):357–380.
- Chantigny, M. H., Angers, D. A., Prévost, D., Simard, R. R., and Chalifour, F.-P. (1999). Dynamics of soluble organic C and C mineralization in cultivated soils with varying N fertilization. *Soil Biology and Biochemistry*, 31(4):543–550.



- Chassé, A. W., Ohno, T., Higgins, S. R., Amirbahman, A., Yildirim, N., and Parr, T. B. (2015). Chemical Force Spectroscopy Evidence Supporting the Layer-by-Layer Model of Organic Matter Binding to Iron (oxy)Hydroxide Mineral Surfaces. *Environmental Science & Technology*, 49(16):9733–9741.
- Chen, J., LeBoeuf, E. J., Dai, S., and Gu, B. (2003). Fluorescence spectroscopic studies of natural organic matter fractions. *Chemosphere*, 50(5):639–647.
- Chen, L., Liu, L., Mao, C., Qin, S., Wang, J., Liu, F., Blagodatsky, S., Yang, G., Zhang, Q., Zhang, D., Yu, J., and Yang, Y. (2018). Nitrogen availability regulates topsoil carbon dynamics after permafrost thaw by altering microbial metabolic efficiency. *Nature Communications*, 9(1):3951.
- Chen, M., Wei, H., Cao, J., Liu, R., Wang, Y., and Zhen, C. (2007). Expression of *Bacillus subtilis proBA* Genes and Reduction of Feedback Inhibition of Proline Synthesis Increases Proline Production and Confers Osmotolerance in Transgenic *Arabidopsis*. *Journal of Biochemistry and Molecular Biology*, 40(3):396–403.
- Cheng, W. and Kuzyakov, Y. (2005). Root effects on soil organic matter decomposition. In Zobel, R. W. and Wright, S. F., editors, *Roots and Soil Management: Interactions between Roots and the Soil*, pages 119–143. Agronomy Monograph 48, American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI, USA.
- Chin, Y.-P., Aiken, G., and O’Loughlin, E. (1994). Molecular Weight, Polydispersity, and Spectroscopic Properties of Aquatic Humic Substances. *Environmental Science & Technology*, 28(11):1853–1858.
- Chow, A. T., Tanji, K. K., Gao, S., and Dahlgren, R. A. (2006). Temperature, water content and wet-dry cycle effects on DOC production and carbon mineralization in agricultural peat soils. *Soil Biology and Biochemistry*, 38(3):477–488.
- Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.
- Coble, P. G. (1996). Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry*, 51(4):325–346.
- Collins, A. N., Sheldrake, G. N., and Crosby, J. (1997). *Chirality in industry II: Developments in the commercial manufacture and applications of optically active compounds*, volume 2. John Wiley & Sons.
- Cook, B. D. and Allan, D. L. (1992). Dissolved organic carbon in old field soils: Total amounts as a measure of available resources for soil mineralization. *Soil Biology and Biochemistry*, 24(6):585–594.

- Cook, F. J. and Knight, J. H. (2003). Oxygen Transport to Plant Roots: Modeling for physical understanding of soil aeration. *Soil Science Society of America Journal*, 67:20–31.
- Cortés-Francisco, N. and Caixach, J. (2013). Molecular Characterization of Dissolved Organic Matter through a Desalination Process by High Resolution Mass Spectrometry. *Environmental Science & Technology*, 47(17):9619–9627.
- Cory, R. M. and McKnight, D. M. (2005). Fluorescence Spectroscopy Reveals Ubiquitous Presence of Oxidized and Reduced Quinones in Dissolved Organic Matter. *Environmental Science & Technology*, 39(21):8142–8149.
- Cosentino, D., Chenu, C., and Le Bissonnais, Y. (2006). Aggregate stability and microbial community dynamics under drying–wetting cycles in a silt loam soil. *Soil Biology and Biochemistry*, 38(8):2053–2062.
- Crameri, F. (2018a). Geodynamic diagnostics, scientific visualisation and StagLab 3.0. *Geoscientific Model Development Discussions*, 11:2541–2562.
- Crameri, F. (2018b). Scientific Colour-Maps: Perceptually uniform and colour-blind friendly.
- Croué, J.-P., Violleau, D., Bodaire, C., and Legube, B. (1999). Removal of Hydrophobic and Hydrophilic Constituents by Anion Exchange Resin. *Water Science and Technology*, 40(9):207–214.
- Csonka, L. N. (1989). Physiological and genetic responses of bacteria to osmotic stress. *Microbiological reviews*, 53(1):121–147.
- Curtin, D., Beare, M. H., and Hernandez-Ramirez, G. (2012). Temperature and Moisture Effects on Microbial Biomass and Soil Organic Matter Mineralization. *Soil Science Society of America Journal*, 76:2055–2067.
- Curtin, D., Beare, M. H., Scott, C. L., Hernandez-Ramirez, G., and Meenken, E. D. (2014). Mineralization of Soil Carbon and Nitrogen Following Physical Disturbance: A Laboratory Assessment. *Soil Science Society of America Journal*, 78:925–935.
- Curtin, D., Wright, C. E., Beare, M. H., and McCallum, F. M. (2006). Hot Water-Extractable Nitrogen as an Indicator of Soil Nitrogen Availability. *Soil Science Society of America Journal*, 70:1512–1521.
- Cuss, C. W. and Guéguen, C. (2015a). Characterizing the Labile Fraction of Dissolved Organic Matter in Leaf Leachates: Methods, Indicators, Structure, and

- Complexity. In *Labile Organic Matter—Chemical Compositions, Function, and Significance in Soil and the Environment*, SSSA Special Publication SV - 62, pages 237–274. Soil Science Society of America, Inc., Madison, WI, USA.
- Cuss, C. W. and Guéguen, C. (2015b). Relationships between molecular weight and fluorescence properties for size-fractionated dissolved organic matter from fresh and aged sources. *Water Research*, 68:487–497.
- Dahlgren, R. A., Percival, H. J., and Parfitt, R. L. (1997). Carbon dioxide degassing effects on soil solutions collected by centrifugation. Technical Report 9, Manaaki Whenua-Landcare Research, Palmerston North, New Zealand.
- Dai, J., Styles, G. N., Patti, A. F., and Saito, K. (2018).  $\text{CuSO}_4/\text{H}_2\text{O}_2$ -Catalyzed Lignin Depolymerization under the Irradiation of Microwaves. *ACS Omega*, 3(9):10433–10441.
- D’Andrilli, J., Cooper, W. T., Foreman, C. M., and Marshall, A. G. (2015). An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. *Rapid communications in mass spectrometry : RCM*, 29(24):2385–2401.
- D’Andrilli, J., Foreman, C. M., Marshall, A. G., and McKnight, D. M. (2013). Characterization of IHSS Pony Lake fulvic acid dissolved organic matter by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and fluorescence spectroscopy. *Organic Geochemistry*, 65:19–28.
- D’Andrilli, J., Junker, J. R., Smith, H. J., Scholl, E. A., and Foreman, C. M. (2019). DOM composition alters ecosystem function during microbial processing of isolated sources. *Biogeochemistry*, 142(2):281–298.
- Davies, C. W. and Thomas, G. G. (1951). Molecular adsorption on ion-exchange resins. *Journal of the Chemical Society (Resumed)*, 0(583):2624–2627.
- De Nobili, M., Contin, M., Mondini, C., and Brookes, P. C. (2001). Soil microbial biomass is triggered into activity by trace amounts of substrate. *Soil Biology and Biochemistry*, 33(9):1163–1170.
- Denef, K., Six, J., Bossuyt, H., Frey, S. D., Elliott, E. T., Merckx, R., and Paustian, K. (2001a). Influence of dry–wet cycles on the interrelationship between aggregate, particulate organic matter, and microbial community dynamics. *Soil Biology and Biochemistry*, 33(12):1599–1611.
- Denef, K., Six, J., Paustian, K., and Merckx, R. (2001b). Importance of macroaggregate dynamics in controlling soil carbon stabilization: short-term effects of physical disturbance induced by dry–wet cycles. *Soil Biology and Biochemistry*, 33(15):2145–2153.

- Desiderio, D. M., editor (1994). *Mass Spectrometry: Clinical and Biomedical Applications*, volume 2. Springer Science + Business Media New York.
- Devey, K., Rajendram, G., Nicholson, B., Hill, R., and Littler, R. (2010). The determination of soil pH – Reducing extraction time from 16 h to 1 h. In Currie, L. D. and Christensen, C. L., editors, *Farming's future: minimising footprints and maximising margins*, Occasional Report 23. Massey University, Palmerston North, New Zealand, Fertilizer and Lime Research Centre.
- Dexter, A. R. (1988). Advances in characterization of soil structure. *Soil and Tillage Research*, 11(3):199–238.
- DiDonato, N., Chen, H., Waggoner, D., and Hatcher, P. G. (2016). Potential origin and formation for molecular components of humic acids in soils. *Geochimica et Cosmochimica Acta*, 178:210–222.
- Dominiak, D. M., Nielsen, J. L., and Nielsen, P. H. (2011). Extracellular DNA is abundant and important for microcolony strength in mixed microbial biofilms. *Environmental Microbiology*, 13(3):710–721.
- Dudal, Y., Sévenier, G., Dupont, L., and Guillon, E. (2005). Fate of the metal-binding soluble organic matter throughout a soil profile. *Soil Science*, 170(9):707–715.
- Dungait, J. A. J., Hopkins, D. W., Gregory, A. S., and Whitmore, A. P. (2012). Soil organic matter turnover is governed by accessibility not recalcitrance. *Global Change Biology*, 18(6):1781–1796.
- Edzwald, J. K. (1993). Coagulation in Drinking Water Treatment: Particles, Organics and Coagulants. *Water Science and Technology*, 27(11):21–35.
- Ekiel, I., Smith, I. C. P., and Sprott, G. D. (1984). Biosynthesis of Isoleucine in Methanogenic Bacteria: A  $^{13}\text{C}$  NMR study. *Biochemistry*, 23(8):1683–1687.
- Eliuk, S. and Makarov, A. (2015). Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual Review of Analytical Chemistry*, 8(1):61–80.
- Embacher, A., Zsolnay, Á., Gattinger, A., and Munch, J. C. (2007). The dynamics of water extractable organic matter (WEOM) in common arable topsoils: I. Quantity, quality and function over a three year period. *Geoderma*, 139(1):11–22.
- Embacher, A., Zsolnay, Á., Gattinger, A., and Munch, J. C. (2008). The dynamics of water extractable organic matter (WEOM) in common arable topsoils: II. Influence of mineral and combined mineral and manure fertilization in a Haplic Chernozem. *Geoderma*, 148(1):63–69.

- Fanin, N., Hättenschwiler, S., and Fromin, N. (2014). Litter fingerprint on microbial biomass, activity, and community structure in the underlying soil. *Plant and Soil*, 379(1-2):79–91.
- Farrell, M., Hill, P. W., Farrar, J., Bardgett, R. D., and Jones, D. L. (2011). Seasonal variation in soluble soil carbon and nitrogen across a grassland productivity gradient. *Soil Biology and Biochemistry*, 43(4):835–844.
- Farrell, M., Hill, P. W., Farrar, J., DeLuca, T. H., Roberts, P., Kielland, K., Dahlgren, R., Murphy, D. V., Hobbs, P. J., Bardgett, R. D., and Jones, D. L. (2013). Oligopeptides Represent a Preferred Source of Organic N Uptake: A Global Phenomenon? *Ecosystems*, 16(1):133–145.
- Farrell, M., Prendergast-Miller, M., Jones, D. L., Hill, P. W., and Condrón, L. M. (2014). Soil microbial organic nitrogen uptake is regulated by carbon availability. *Soil Biology and Biochemistry*, 77:261–267.
- Favoretto, C. M., Gonçalves, D., Milori, D. M. B. P., Rosa, J. A., Leite, W. C., Brinatti, A. M., and Saab, S. d. C. (2008). Determinação da humificação da matéria orgânica de um latossolo e de suas frações organo-minerais. *Química Nova*, 31(8):1994–1996.
- Fellman, J. B., D’Amore, D. V., Hood, E., and Boone, R. D. (2008). Fluorescence characteristics and biodegradability of dissolved organic matter in forest and wetland soils from coastal temperate watersheds in southeast Alaska. *Biogeochemistry*, 88(2):169–184.
- Fellman, J. B., Hood, E., Edwards, R. T., and D’Amore, D. V. (2009). Changes in the concentration, biodegradability, and fluorescent properties of dissolved organic matter during stormflows in coastal temperate watersheds. *Journal of Geophysical Research: Biogeosciences*, 114(G01021).
- Fellman, J. B., Hood, E., and Spencer, R. G. M. (2010). Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnology and Oceanography*, 55(6):2452–2462.
- Fernández-Romero, M. L., Clark, J. M., Collins, C. D., Parras-Alcántara, L., and Lozano-García, B. (2016). Evaluation of optical techniques for characterising soil organic matter quality in agricultural soils. *Soil and Tillage Research*, 155:450–460.
- Fierer, N. and Schimel, J. P. (2002). Effects of drying–rewetting frequency on soil carbon and nitrogen transformations. *Soil Biology and Biochemistry*, 34(6):777–787.

- Fierer, N., Schimel, J. P., and Holden, P. A. (2003). Influence of Drying–Rewetting Frequency on Soil Bacterial Community Structure. *Microbial Ecology*, 45(1):63–71.
- Figueiredo, M. V. B., Burity, H. A., Martínez, C. R., and Chanway, C. P. (2008). Alleviation of drought stress in the common bean (*Phaseolus vulgaris* L.) by co-inoculation with *Paenibacillus polymyxa* and *Rhizobium tropici*. *Applied Soil Ecology*, 40(1):182–188.
- Filley, T. R., Boutton, T. W., Liao, J. D., Jastrow, J. D., and Gamblin, D. E. (2008). Chemical changes to nonaggregated particulate soil organic matter following grassland-to-woodland transition in a subtropical savanna. *Journal of Geophysical Research: Biogeosciences*, 113(G03009).
- Flemming, H.-C., Neu, T. R., and Wozniak, D. J. (2007). The EPS Matrix: The “House of Biofilm Cells”. *Journal of Bacteriology*, 189(22):7945–7947.
- Fontaine, S., Henault, C., Aamor, A., Bdioui, N., Bloor, J. M. G., Maire, V., Mary, B., Revalliot, S., and Maron, P. A. (2011). Fungi mediate long term sequestration of carbon and nitrogen in soil through their priming effect. *Soil Biology and Biochemistry*, 43(1):86–96.
- Francis, G. S., Tabley, F. J., and White, K. M. (1999). Restorative crops for the amelioration of degraded soil conditions in New Zealand. *Soil Research*, 37(6):1017–1034.
- Frank, E., Hall, M., and Pfahringer, B. (2003). Locally Weighted Naive Bayes. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, UAI’03, pages 249–256, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Frank, E., Hall, M. A., and Witten, I. H. (2016). The WEKA workbench. In Kaufmann, M., editor, *Data Mining: Practical Machine Learning Tools and Techniques*, 4<sup>th</sup> edition, page Online Appendix.
- Fraser, P. M., Curtin, D., Harrison-Kirk, T., Meenken, E. D., Beare, M. H., Tabley, F., Gillespie, R. N., and Francis, G. S. (2013). Winter Nitrate Leaching under Different Tillage and Winter Cover Crop Management Practices. *Soil Science Society of America Journal*, 77:1391–1401.
- Fröberg, M., Kleja, D. B., Bergkvist, B., Tipping, E., and Mulder, J. (2005). Dissolved organic carbon leaching from a coniferous forest floor—a field manipulation experiment. *Biogeochemistry*, 75(2):271–287.

- Fu, P. L. K. and Symons, J. M. (1990). Removing Aquatic Organic Substances by Anion Exchange Resins. *Journal - American Water Works Association*, 82(10):70–77.
- Fuchs, G., Boll, M., and Heider, J. (2011). Microbial degradation of aromatic compounds – from one strategy to four. *Nature Reviews Microbiology*, 9:803.
- Fuentes, M., González-Gaitano, G., and García-Mina, J. M. (2006). The usefulness of UV–visible and fluorescence spectroscopies to study the chemical nature of humic substances from soils and composts. *Organic Geochemistry*, 37(12):1949–1959.
- Gan, C. S., Chong, P. K., Pham, T. K., and Wright, P. C. (2007). Technical, Experimental, and Biological Variations in Isobaric Tags for Relative and Absolute Quantitation (iTRAQ). *Journal of Proteome Research*, 6(2):821–827.
- Gao, H., Chen, X., Wei, J., Zhang, Y., Zhang, L., Chang, J., and Thompson, M. L. (2016a). Decomposition Dynamics and Changes in Chemical Composition of Wheat Straw Residue under Anaerobic and Aerobic Conditions. *PLoS ONE*, 11(7):e0158172–e0158172.
- Gao, J., Feng, J., Zhang, X., Yu, F.-H., Xu, X., and Kuzyakov, Y. (2016b). Drying–rewetting cycles alter carbon and nitrogen mineralization in litter-amended alpine wetland soil. *CATENA*, 145:285–290.
- Gärdenäs, A. I., Ågren, G. I., Bird, J. A., Clarholm, M., Hallin, S., Ineson, P., Kätterer, T., Knicker, H., Nilsson, S. I., Näsholm, T., Ogle, S., Paustian, K., Persson, T., and Stendahl, J. (2011). Knowledge gaps in soil carbon and nitrogen interactions – From molecular to global scale. *Soil Biology and Biochemistry*, 43(4):702–717.
- Gašparović, B., Penezić, A., Lampitt, R. S., Sudasinghe, N., and Schaub, T. (2017). Depth-related cycling of suspended nitrogen-containing lipids in the northeast Atlantic. *Organic Geochemistry*, 113:55–66.
- Gee, G. W. and Or, D. (2002). 2.4 Particle-Size Analysis. In *Methods of Soil Analysis: Part 4 Physical Methods*, SSSA Book Series SV - 5.4, pages 255–293. Soil Science Society of America, Madison, WI.
- Geisseler, D., Horwath, W. R., Joergensen, R. G., and Ludwig, B. (2010). Pathways of nitrogen utilization by soil microorganisms – A review. *Soil Biology and Biochemistry*, 42(12):2058–2067.

- Ghani, A., Dexter, M., Carran, R. A., and Theobald, P. W. (2007). Dissolved organic nitrogen and carbon in pastoral soils: the New Zealand experience. *European Journal of Soil Science*, 58(3):832–843.
- Ghani, A., Dexter, M., and Perrott, K. W. (2003). Hot-water extractable carbon in soils: a sensitive measurement for determining impacts of fertilisation, grazing and cultivation. *Soil Biology and Biochemistry*, 35(9):1231–1243.
- Ghani, A., Müller, K., Dodd, M., and Mackay, A. (2010). Dissolved organic matter leaching in some contrasting New Zealand pasture soils. *European Journal of Soil Science*, 61(4):525–538.
- Ghasemi, J. and Tavakoli, H. (2013). Application of Random Forest Regression to Spectral Multivariate Calibration. *Analytical Methods*, 5:1863–1871.
- Ghosh, K. and Schnitzer, M. (1980). Fluorescence excitation spectra of humic substances. *Canadian Journal of Soil Science*, 60(2):373–379.
- Gilcrease, P. C. and Murphy, V. G. (1995). Bioconversion of 2,4-diamino-6-nitrotoluene to a novel metabolite under anoxic and aerobic conditions. *Applied and Environmental Microbiology*, 61(12):4209–4214.
- Gilli, G. and Gilli, P. (2009). *The Nature of the Hydrogen Bond: Outline of a Comprehensive Hydrogen Bond Theory*. Oxford University Press, Oxford, UK.
- Glick, D. (2009). *Methods of biochemical analysis*, volume 25. John Wiley & Sons, New York, USA.
- Gloux, K. and Le Rudulier, D. (1989). Transport and catabolism of proline betaine in salt-stressed *Rhizobium meliloti*. *Archives of Microbiology*, 151(2):143–148.
- Golchin, A., Oades, J. M., Skjemstad, J. O., and Clarke, P. (1994). Soil structure and carbon cycling. *Soil Research*, 32(5):1043–1068.
- Goldmann, A., Boivin, C., Fleury, V., Message, B., Lecoœur, L., Maille, M., and Tepfer, D. (1991). Betaine use by rhizosphere bacteria: genes essential for trigonelline, stachydrine, and carnitine catabolism in *Rhizobium meliloti* are located on pSym in the symbiotic region. *Molecular Plant-Microbe Interactions: MPMI*, 4(6):571–578.
- González-Pérez, M., Milori, D. M. B. P., Colnago, L. A., Martin-Neto, L., and Melo, W. J. (2007). A laser-induced fluorescence spectroscopic study of organic matter in a Brazilian Oxisol under different tillage systems. *Geoderma*, 138(1):20–24.



- Grandy, A. S. and Robertson, G. P. (2006). Aggregation and Organic Matter Protection Following Tillage of a Previously Uncultivated Soil. *Soil Science Society of America Journal*, 70:1398–1406.
- Green, S. A. and Blough, N. V. (1994). Optical absorption and fluorescence properties of chromophoric dissolved organic matter in natural waters. *Limnology and Oceanography*, 39(8):1903–1916.
- Gregorich, E. G., Beare, M. H., Stoklas, U., and St-Georges, P. (2003). Biodegradability of soluble organic matter in maize-cropped soils. *Geoderma*, 113(3):237–252.
- Gregorich, E. G., Gillespie, A. W., Beare, M. H., Curtin, D., Sanei, H., and Yanni, S. F. (2015). Evaluating biodegradability of soil organic matter by its thermal stability and chemical composition. *Soil Biology and Biochemistry*, 91:182–191.
- Gregorich, E. G., Liang, B. C., Drury, C. F., Mackenzie, A. F., and McGill, W. B. (2000). Elucidation of the source and turnover of water soluble and microbial biomass carbon in agricultural soils. *Soil Biology and Biochemistry*, 32(5):581–587.
- Griffin, B. A. and Jurinak, J. J. (1973). Estimation of activity coefficients from the electrical conductivity of natural aquatic systems and soil extracts. *Soil Science*, 116(1).
- Guggenberger, G., Christensen, B. T., and Zech, W. (1994a). Land-use effects on the composition of organic matter in particle-size separates of soil: I. Lignin and carbohydrate signature. *European Journal of Soil Science*, 45(4):449–458.
- Guggenberger, G. and Zech, W. (1994). Composition and dynamics of dissolved carbohydrates and lignin-degradation products in two coniferous forests, N.E. Bavaria, Germany. *Soil Biology and Biochemistry*, 26(1):19–27.
- Guggenberger, G., Zech, W., Haumeier, L., and Christensen, B. T. (1995). Land-use effects on the composition of organic matter in particle-size separates of soils: II. CPMAS and solution  $^{13}\text{C}$  NMR analysis. *European Journal of Soil Science*, 46(1):147–158.
- Guggenberger, G., Zech, W., and Schulten, H.-R. (1994b). Formation and mobilization pathways of dissolved organic matter: evidence from chemical structural studies of organic matter fractions in acid forest floor solutions. *Organic Geochemistry*, 21(1):51–66.

- Guigue, J., Mathieu, O., Lévêque, J., Mounier, S., Laffont, R., Maron, P. A., Navarro, N., Chateau, C., Amiotte-Suchet, P., and Lucas, Y. (2014). A comparison of extraction procedures for water-extractable organic matter in soils. *European Journal of Soil Science*, 65(4):520–530.
- Guo, L. B. and Gifford, R. M. (2002). Soil carbon stocks and land use change: a meta analysis. *Global Change Biology*, 8(4):345–360.
- Hadley, W. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Halket, J. M., Waterman, D., Przyborowska, A. M., Patel, R. K. P., Fraser, P. D., and Bramley, P. M. (2005). Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *Journal of Experimental Botany*, 56(410):219–243.
- Halverson, L. J., Jones, T. M., and Firestone, M. K. (2000). Release of Intracellular Solutes by Four Soil Bacteria Exposed to Dilution Stress. *Soil Science Society of America Journal*, 64:1630–1637.
- Hamilton III, E. W. and Frank, D. A. (2001). Can plants stimulate soil microbes and their own nutrient supply? Evidence from a grazing tolerant grass. *Ecology*, 82(9):2397–2402.
- Hansen, A. M., Kraus, T. E. C., Pellerin, B. A., Fleck, J. A., Downing, B. D., and Bergamaschi, B. A. (2016). Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation. *Limnology and Oceanography*, 61(3):1015–1032.
- Harris, R. (1981). Effect of water potential on microbial growth and activity. In Parr, J. F., Gardner, W. R., and Elliott, L. F., editors, *Water Potential Relations in Soil Microbiology*, pages 23–96. SSSA Spec. Publ. 9, Madison, WI, USA.
- Harrison-Kirk, T., Beare, M. H., Meenken, E. D., and Condrón, L. M. (2013). Soil organic matter and texture affect responses to dry/wet cycles: Effects on carbon dioxide and nitrous oxide emissions. *Soil Biology and Biochemistry*, 57:43–55.
- Harrison-Kirk, T., Beare, M. H., Meenken, E. D., and Condrón, L. M. (2014). Soil organic matter and texture affect responses to dry/wet cycles: Changes in soil organic matter fractions and relationships with C and N mineralisation. *Soil Biology and Biochemistry*, 74:50–60.
- Harshman, R. (1970). Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84.

- Hartley, I. P. and Ineson, P. (2008). Substrate quality and the temperature sensitivity of soil organic matter decomposition. *Soil Biology and Biochemistry*, 40(7):1567–1574.
- Harwood, C. S., Burchhardt, G., Herrmann, H., and Fuchs, G. (1998). Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. *FEMS Microbiology Reviews*, 22(5):439–458.
- Haumaier, L. and Zech, W. (1995). Black carbon—possible source of highly aromatic components of soil humic acids. *Organic Geochemistry*, 23(3):191–196.
- Haynes, R. J. (2000). Labile organic matter as an indicator of organic matter quality in arable and pastoral soils in New Zealand. *Soil Biology and Biochemistry*, 32(2):211–219.
- Haynes, R. J. (2005). Labile Organic Matter Fractions as Central Components of the Quality of Agricultural Soils: An Overview. *Advances in Agronomy*, 85:221–268.
- Haynes, R. J. and Francis, G. S. (1993). Changes in microbial biomass C, soil carbohydrate composition and aggregate stability induced by growth of selected crop and forage species under field conditions. *Journal of Soil Science*, 44(4):665–675.
- Haynes, R. J. and Swift, R. S. (1990). Stability of soil aggregates in relation to organic constituents and soil water content. *Journal of Soil Science*, 41(1):73–83.
- Hedges, J. I. and Ertel, J. R. (1982). Characterization of lignin by gas capillary chromatography of cupric oxide oxidation products. *Analytical Chemistry*, 54(2):174–178.
- Helms, J. R., Stubbins, A., Ritchie, J. D., Minor, E. C., Kieber, D. J., and Mopper, K. (2008). Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography*, 53(3):955–969.
- Hemström, P. and Irgum, K. (2006). Hydrophilic interaction chromatography. *Journal of Separation Science*, 29(12):1784–1821.
- Hermesen, R., Okano, H., You, C., Werner, N., and Hwa, T. (2015). A growth-rate composition formula for the growth of *E.coli* on co-utilized carbon substrates. *Molecular Systems Biology*, 11(4):801.

- Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Kopplin, P., Witt, M., Kaiser, K., Kettrup, A., and Hedges, J. I. (2006). Characterization of a major refractory component of marine dissolved organic matter. *Geochimica et Cosmochimica Acta*, 70(12):2990–3010.
- Hewitt, A. E. (2010). New Zealand soil classification. Technical Report No.1, Lincoln, New Zealand.
- Higgins, C. F. and Payne, J. W. (1980). Transport and utilization of amino acids and peptides by higher plants. In Payne, J. W., editor, *Microorganisms and nitrogen sources*, pages 609–641. John Wiley & Sons, Chichester, UK.
- Hill, P. W., Farrell, M., and Jones, D. L. (2012). Bigger may be better in soil N cycling: Does rapid acquisition of small L-peptides by soil microbes dominate fluxes of protein-derived N in soil? *Soil Biology and Biochemistry*, 48:106–112.
- Hill, P. W., Quilliam, R. S., DeLuca, T. H., Farrar, J., Farrell, M., Roberts, P., Newsham, K. K., Hopkins, D. W., Bardgett, R. D., and Jones, D. L. (2011). Acquisition and assimilation of nitrogen as peptide-bound and D-enantiomers of amino acids by wheat. *PLoS ONE*, 6(4):e19220.
- Hockaday, W. C., Grannas, A. M., Kim, S., and Hatcher, P. G. (2006). Direct molecular evidence for the degradation and mobility of black carbon in soils from ultrahigh-resolution mass spectral analysis of dissolved organic matter from a fire-impacted forest soil. *Organic Geochemistry*, 37(4):501–510.
- Hodson, M., Vijver, M., and J.G.M. Peijnenburg, W. (2011). Bioavailability in soils. In Swartjes, F. A., editor, *Dealing with contaminated sites: from theory towards practical application*, pages 721–746. Springer.
- Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.
- Hopkins, D. W. (2008). Carbon mineralization. In *Soil Sampling and Methods of Analysis*, pages 589–598. Canadian Society of Soil Science Publication, CRC Press Boca Raton.
- Horwath, W. (2015). Carbon cycling: The dynamics and formation of organic matter. In Paul, E. A., editor, *Soil Microbiology Ecology, and Biochemistry*, pages 339–382. Academic Press, 4<sup>th</sup> edition.
- Howard, D. M. and Howard, P. J. A. (1993). Relationships between CO<sub>2</sub> evolution, moisture content and temperature for a range of soil types. *Soil Biology and Biochemistry*, 25(11):1537–1546.

- Howell, D. M., Xu, H., and White, R. H. (1999). (*R*)-Citramalate Synthase in Methanogenic Archaea. *Journal of Bacteriology*, 181(1):331–333.
- Hsu, P.-H. and Hatcher, P. G. (2005). New evidence for covalent coupling of peptides to humic acids based on 2D NMR spectroscopy: A means for preservation. *Geochimica et Cosmochimica Acta*, 69(18):4521–4533.
- Huang, P. M., Li, Y., and Sumner, M. E. (2012). *Handbook of soil sciences: properties and processes*. CRC Press, Boca Raton, FL, USA, 2<sup>nd</sup> edition.
- Hudson, N., Baker, A., and Reynolds, D. (2007). Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters – A review. *River Research and Applications*, 23:631–649.
- Inamdar, S., Finger, N., Singh, S., Mitchell, M., Levina, D., Bais, H., Scott, D., and McHale, P. (2012). Dissolved organic matter (DOM) concentration and quality in a forested mid-Atlantic watershed, USA. *Biogeochemistry*, 108(1):55–76.
- Ioannidis, S. and Anderko, A. (2001). Equilibrium Modeling of Combined Ion-Exchange and Molecular Adsorption Phenomena. *Industrial & Engineering Chemistry Research*, 40(2):714–720.
- IUSS Working Group WRB (2015). *World Reference Base for Soil Resources 2014, update 2015 International soil classification system for naming soils and creating legends for soil maps*. World Soil Resources Reports No. 106. FAO, Rome.
- Iversen, G. R. (2011). Analysis of Variance. In Lovric, M., editor, *International Encyclopedia of Statistical Science*, pages 44–46. Springer-Verlag Berlin Heidelberg, 1<sup>st</sup> edition.
- Jaffé, R., McKnight, D., Maie, N., Cory, R., McDowell, W. H., and Campbell, J. L. (2008). Spatial and temporal variations in DOM composition in ecosystems: The importance of long-term monitoring of optical properties. *Journal of Geophysical Research: Biogeosciences*, 113(G04032).
- Jaffrain, J., Gérard, F., Meyer, M., and Ranger, J. (2007). Assessing the Quality of Dissolved Organic Matter in Forest Soils Using Ultraviolet Absorption Spectrophotometry. *Soil Science Society of America Journal*, 71:1851–1858.
- Jolliffe, I. (2011). Principal Component Analysis. In Lovric, M., editor, *International Encyclopedia of Statistical Science*, pages 1094–1096. Springer-Verlag Berlin Heidelberg, 1<sup>st</sup> edition.
- Jones, D. L. and Murphy, D. V. (2007). Microbial response time to sugar and amino acid additions to soil. *Soil Biology and Biochemistry*, 39(8):2178–2182.

- Jongmans, A. G., Pulleman, M. M., and Marinissen, J. C. Y. (2001). Soil structure and earthworm activity in a marine silt loam under pasture versus arable land. *Biology and Fertility of Soils*, 33(4):279–285.
- Kaiser, K., Kaupenjohann, M., and Zech, W. (2001). Sorption of dissolved organic carbon in soils: effects of soil sample storage, soil-to-solution ratio, and temperature. *Geoderma*, 99(3):317–328.
- Kaiser, K. and Zech, W. (1997). Competitive sorption of dissolved organic matter fractions to soils and related mineral phases. *Soil Science Society of America Journal*, 61(1):64–69.
- Kaiser, M., Kleber, M., and Berhe, A. A. (2015). How air-drying and rewetting modify soil organic matter characteristics: An assessment to improve data interpretation and inference. *Soil Biology and Biochemistry*, 80:324–340.
- Kalbitz, K., Geyer, W., and Geyer, S. (1999). Spectroscopic properties of dissolved humic substances – a reflection of land use history in a fen area. *Biogeochemistry*, 47(2):219–238.
- Kalbitz, K., Schmerwitz, J., Schwesig, D., and Matzner, E. (2003). Biodegradation of soil-derived dissolved organic matter as related to its properties. *Geoderma*, 113(3):273–291.
- Kalbitz, K., Solinger, S., Park, J.-H., Michalzik, B., and Matzner, E. (2000). Controls on the dynamics of dissolved organic matter in soils: a review. *Soil Science*, 165(4).
- Kallenbach, C. M., Frey, S. D., and Grandy, A. S. (2016). Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Nature Communications*, 7(13630):1–10.
- Kanazawa, N., Urano, K., Kokado, N., and Urushigawa, Y. (2001). Adsorption Equilibrium Equation of Carboxylic Acids on Anion-Exchange Resins in Water. *Journal of Colloid and Interface Science*, 238(1):196–202.
- Kanazawa, N., Urano, K., Kokado, N., and Urushigawa, Y. (2004). Exchange characteristics of monocarboxylic acids and monosulfonic acids onto anion-exchange resins. *Journal of Colloid and Interface Science*, 271(1):20–27.
- Karavanova, E. and Milanovskiy, E. (2016). Aromaticity and humification of dissolved organic matter (lysimetric experiment). In *International Soil Science Congress on “Soil Science in International Year of Soils 2015”*, pages 178–182.

- Kätterer, T., Bolinder, M. A., Andrén, O., Kirchmann, H., and Menichetti, L. (2011). Roots contribute more to refractory soil organic matter than above-ground crop residues, as revealed by a long-term field experiment. *Agriculture, Ecosystems & Environment*, 141(1):184–192.
- Keeney, D. R. and Nelson, D. W. (1982). Nitrogen–Inorganic Forms. In *Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties*, Agronomy Monograph SV - 9.2, pages 643–698. American Society of Agronomy, Soil Science Society of America, Madison, WI, USA.
- Kemmitt, S. J., Lanyon, C. V., Waite, I. S., Wen, Q., Addiscott, T. M., Bird, N. R. A., O'donnell, A. G., and Brookes, P. C. (2008). Mineralization of native soil organic matter is not regulated by the size, activity or composition of the soil microbial biomass—a new perspective. *Soil Biology and Biochemistry*, 40(1):61–73.
- Kemper, W., Rosenau, R., and Dexter, A. (1987). Cohesion Development in Disrupted Soils as Affected by Clay and Organic Matter Content and Temperature. *Soil Science Society of America Journal*, 51(4):860–867.
- Kempf, B. and Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Archives of Microbiology*, 170(5):319–330.
- Kidd, D. R., Ryan, M. H., Hahne, D., Haling, R. E., Lambers, H., Sandral, G. A., Simpson, R. J., and Cawthray, G. R. (2018). The carboxylate composition of rhizosheath and root exudates from twelve species of grassland and crop legumes with special reference to the occurrence of citramalate. *Plant and Soil*, 424(1):389–403.
- Kidd, D. R., Ryan, M. H., Haling, R. E., Lambers, H., Sandral, G. A., Yang, Z., Culvenor, R. A., Cawthray, G. R., Stefanski, A., and Simpson, R. J. (2016). Rhizosphere carboxylates and morphological root traits in pasture legumes and grasses. *Plant and Soil*, 402(1):77–89.
- Kieft, T. L., Soroker, E., and Firestone, M. K. (1987). Microbial biomass response to a rapid increase in water potential when dry soil is wetted. *Soil Biology and Biochemistry*, 19(2):119–126.
- Kiem, R. and Kögel-Knabner, I. (2003). Contribution of lignin and polysaccharides to the refractory carbon pool in C-depleted arable soils. *Soil Biology and Biochemistry*, 35(1):101–118.

- Kim, H.-C., Yu, M.-J., and Han, I. (2006). Multi-method study of the characteristic chemical nature of aquatic humic substances isolated from the Han River, Korea. *Applied Geochemistry*, 21(7):1226–1239.
- Kim, S., Kramer, R. W., and Hatcher, P. G. (2003). Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Analytical Chemistry*, 75(20):5336–5344.
- Kind, T. and Fiehn, O. (2006). Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC bioinformatics*, 7:234.
- Kind, T. and Fiehn, O. (2007). Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC bioinformatics*, 8:105.
- Kingdon, K. H. (1923). A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Physical Review*, 21:408–418.
- Kleber, M., Nico, P. S., Plante, A., Filley, T., Kramer, M., Swanston, C., and Sollins, P. (2011). Old and stable soil organic matter is not necessarily chemically recalcitrant: implications for modeling concepts and temperature sensitivity. *Global Change Biology*, 17(2):1097–1107.
- Kleber, M., Sollins, P., and Sutton, R. (2007). A Conceptual Model of Organo-Mineral Interactions in Soils: Self-Assembly of Organic Molecular Fragments Into Zonal Structures on Mineral Surfaces. *Biogeochemistry*, 85(1):9–24.
- Klotzbücher, T., Kalbitz, K., Cerli, C., Hernes, P. J., and Kaiser, K. (2016). Gone or just out of sight? The apparent disappearance of aromatic litter components in soils. *SOIL*, 2(3):325–335.
- Klumpp, K., Fontaine, S., Attard, E., Le Roux, X., Gleixner, G., and Soussana, J.-F. (2009). Grazing triggers soil carbon loss by altering plant roots and their control on soil microbial community. *Journal of Ecology*, 97(5):876–885.
- Knicker, H. (2011). Soil organic N – An under-rated player for C sequestration in soils? *Soil Biology and Biochemistry*, 43(6):1118–1129.
- Koch, B. P. and Dittmar, T. (2006). From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry*, 20(5):926–932.



- Koch, B. P. and Dittmar, T. (2016). From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry*, 30(1):250.
- Koch, B. P., Dittmar, T., Witt, M., and Kattner, G. (2007). Fundamentals of Molecular Formula Assignment to Ultrahigh Resolution Mass Data of Natural Organic Matter. *Analytical Chemistry*, 79(4):1758–1763.
- Koch, B. P., Witt, M., Engbrodt, R., Dittmar, T., and Kattner, G. (2005). Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochimica et Cosmochimica Acta*, 69(13):3299–3308.
- Kögel-Knabner, I. (2002). The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. *Soil Biology and Biochemistry*, 34(2):139–162.
- Kögel-Knabner, I. (2006). Chemical Structure of Organic N and Organic P in Soil. In Nannipieri, P. and Smalla, K., editors, *Nucleic Acids and Proteins in Soil*, pages 23–48. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Kögel-Knabner, I., Zech, W., and Hatcher, P. G. (1991). Chemical Structural Studies of Forest Soil Humic Acids: Aromatic Carbon Fraction. *Soil Science Society of America Journal*, 55:241–247.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai '95 - Proceedings of the 14<sup>th</sup> International Joint Conference on Artificial Intelligence*, volume 2, pages 1137–1143. Montreal, Canada.
- Kong, A. Y. Y. and Six, J. (2010). Tracing root vs. residue carbon into soils from conventional and alternative cropping systems. *Soil Science Society of America Journal*, 74(4):1201–1210.
- Kononova, M. M. (1966). *Soil Organic Matter: Its Nature, Its Role in Soil Formation and in Soil Fertility*. Pergamon.
- Korak, J. A., Dotson, A. D., Summers, R. S., and Rosario-Ortiz, F. L. (2014). Critical analysis of commonly used fluorescence metrics to characterize dissolved organic matter. *Water Research*, 49:327–338.
- Korshin, G. V., Li, C.-W., and Benjamin, M. M. (1997). Monitoring the properties of natural organic matter through UV spectroscopy: A consistent theory. *Water Research*, 31(7):1787–1795.

- Kowalenko, C. G., Ivarson, K. C., and Cameron, D. R. (1978). Effect of moisture content, temperature and nitrogen fertilization on carbon dioxide evolution from field soils. *Soil Biology and Biochemistry*, 10(5):417–423.
- Ku, Y., Kuen-Chyr, L., and Wang, W. (2004). Removal of Phenols from Aqueous Solutions by Purolite A-510 Resin. *Separation Science and Technology*, 39(4):911–923.
- Kuhn, M. (2008). Building Predictive Models in R Using the `caret` Package. *Journal of Statistical Software*, 1(5).
- Kuhn, M. (2015). `caret`: Classification and regression training.
- Kuhn, S., Egert, B., Neumann, S., and Steinbeck, C. (2008). Building blocks for automated elucidation of metabolites: machine learning methods for NMR prediction. *BMC bioinformatics*, 9:400.
- Kumar, K. (2019). Optimizing Parallel Factor (PARAFAC) Assisted Excitation-Emission Matrix Fluorescence (EEMF) Spectroscopic Analysis of Multifluorophoric Mixtures. *Journal of Fluorescence*.
- Kumar, K. and Mishra, A. K. (2013). Analysis of dilute aqueous multifluorophoric mixtures using excitation–emission matrix fluorescence (EEMF) and total synchronous fluorescence (TSF) spectroscopy: A comparative evaluation. *Talanta*, 117:209–220.
- Kunhi Mouvenchery, Y., Kučerík, J., Diehl, D., and Schaumann, G. E. (2012). Cation-mediated cross-linking in natural organic matter: a review. *Reviews in Environmental Science and Bio/Technology*, 11(1):41–54.
- Kunin, R. and Suffet, I. H. (1980). Removal of humic material from drinking water by anion exchange resins. In McGuire, M. and Suffet, I., editors, *Activated carbon adsorption of organics from the aqueous phase*, pages 425–442. Ann Arbor Science Publishers Inc, Ann Arbor, MI.
- Kuzyakov, Y., Apostel, C., Gunina, A., Herrmann, A. M., and Dippold, M. (2015). Oxidation state, bioavailability & biochemical pathway define the fate of carbon in soil. In *EGU General Assembly Conference Abstracts*, number EGU2015–3533. EGU General Assembly 2015.
- Kuzyakov, Y., Friedel, J. K., and Stahr, K. (2000). Review of mechanisms and quantification of priming effects. *Soil Biology and Biochemistry*, 32(11):1485–1498.

- Lagomarsino, D., Tofani, V., Segoni, S., Catani, F., and Casagli, N. (2017). A Tool for Classification and Regression Using Random Forest Methodology: Applications to Landslide Susceptibility Mapping and Soil Thickness Modeling. *Environmental Modeling & Assessment*, 22(3):201–214.
- Laird, N. M. and Ware, J. H. (1982). Random-effects Models for Longitudinal Data. *Biometrics*, 38:963–974.
- Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy*. Springer US, 3<sup>rd</sup> edition.
- Lal, R. (2004). Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. *Science*, 304(5677):1623–1627.
- Lam, L. K., Arnold, L. D., Kalantar, T. H., Kelland, J. G., Lane-Bell, P. M., Palcic, M. M., Pickard, M. A., and Vederas, J. C. (1988). Analogs of diaminopimelic acid as inhibitors of meso-diaminopimelate dehydrogenase and LL-diaminopimelate epimerase. *Journal of Biological Chemistry*, 263(24):11814–11819.
- Lazrek, H. B., Rochdi, A., Khaider, H., Barascut, J.-L., Imbach, J.-L., Balzarini, J., Witvrouw, M., Pannecouque, C., and De Clercq, E. (1998). Synthesis of (*Z*) and (*E*)  $\alpha$ -alkenyl phosphonic acid derivatives of purines and pyrimidines. *Tetrahedron*, 54(15):3807–3816.
- Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneeth, A., Arora, V. K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., and Zheng, B. (2018). Global carbon budget 2018. *Earth System Science Data*, 10(4):2141–2194.
- Leenheer, J. A. and Croué, J.-P. (2003). Peer Reviewed: Characterizing Aquatic Dissolved Organic Matter. *Environmental Science & Technology*, 37(1):18A–26A.

- Leenheer, J. A., Nanny, M. A., and McIntyre, C. (2003). Terpenoids as Major Precursors of Dissolved Organic Matter in Landfill Leachates, Surface Water, and Groundwater. *Environmental Science & Technology*, 37(11):2323–2331.
- Lehninger, A. L., Nelson, D. L., and Cox, M. M. (2008). *Lehninger Principles of Biochemistry*. W. H. Freeman, New York, 5<sup>th</sup> edition.
- Leifeld, J. and Kögel-Knabner, I. (2005). Soil organic matter fractions as early indicators for carbon stock changes under different land-use? *Geoderma*, 124(1):143–155.
- Leinweber, P., Schulten, H. R., and Körschens, M. (1995). Hot water extracted organic matter: chemical composition and temporal variations in a long-term field experiment. *Biology and Fertility of Soils*, 20(1):17–23.
- Leurgans, S. and Ross, R. T. (1992). Multilinear Models: Applications in Spectroscopy. *Statistical Science*, 7(3):289–310.
- Li, P. and SenGupta, A. K. (2004). Sorption of hydrophobic ionizable organic compounds (HIOCs) onto polymeric ion exchangers. *Reactive and Functional Polymers*, 60:27–39.
- Li, Z., Zhao, B., Wang, Q., Cao, X., and Zhang, J. (2015). Differences in Chemical Composition of Soil Organic Carbon Resulting From Long-Term Fertilization Strategies. *PLoS ONE*, 10(4):e0124359.
- Liaw, A. and Wiener, M. (2002). Classification and Regression by `randomForest`. *R news*, 2(3):18–22.
- Lichtfouse, É., Berthier, G., Houot, S., Barriuso, E., Bergheaud, V., and Vallaëys, T. (1995). Stable carbon isotope evidence for the microbial origin of C<sub>14</sub>–C<sub>18</sub> n-alkanoic acids in soils. *Organic Geochemistry*, 23(9):849–852.
- Liljeroth, E., Kuikman, P., and Van Veen, J. A. (1994). Carbon translocation to the rhizosphere of maize and wheat and influence on the turnover of native soil organic matter at different soil nitrogen levels. *Plant and Soil*, 161(2):233–240.
- Linn, D. M. and Doran, J. W. (1984). Effect of Water-Filled Pore Space on Carbon Dioxide and Nitrous Oxide Production in Tilled and Nontilled Soils. *Soil Science Society of America Journal*, 48:1267–1272.
- Lippert, K. and Galinski, E. A. (1992). Enzyme stabilization by ectoine-type compatible solutes: protection against heating, freezing and drying. *Applied Microbiology and Biotechnology*, 37(1):61–65.

- Long, F. A. and McDevit, W. F. (1952). Activity Coefficients of Nonelectrolyte Solutes in Aqueous Salt Solutions. *Chemical Reviews*, 51(1):119–169.
- Ludwig, B., Meiwes, K. J., Khanna, P., Gehlen, R., Fortmann, H., and Hildebrand, E. E. (1999). Comparison of different laboratory methods with lysimetry for soil solution composition – experimental and model results. *Journal of Plant Nutrition and Soil Science*, 162(3):343–351.
- Lundquist, E. J., Jackson, L. E., and Scow, K. M. (1999). Wet–dry cycles affect dissolved organic carbon in two California agricultural soils. *Soil Biology and Biochemistry*, 31(7):1031–1038.
- Ly, Q. V. and Hur, J. (2018). Further insight into the roles of the chemical composition of dissolved organic matter (DOM) on ultrafiltration membranes as revealed by multiple advanced DOM characterization tools. *Chemosphere*, 201:168–177.
- MacKay, D. J. C. (1998). Introduction to Gaussian processes. *NATO ASI Series F Computer and Systems Sciences*, 168:133–166.
- Maes, E., Valkenburg, D., Baggerman, G., Willems, H., Landuyt, B., Schoofs, L., and Mertens, I. (2015). Determination of variation parameters as a crucial step in designing TMT-based clinical proteomics experiments. *PLoS ONE*, 10(3):e0120115–e0120115.
- Malcolm, R. L. (1985). The geochemistry of stream fulvic and humic acids. In Aiken, G. R., McKnight, D. M., Wershaw, R. L., and MacCarthy, P., editors, *Humic substances in soil, sediment, and water: geochemistry, isolation and characterization*, pages 181–209. John Wiley & Sons, New York, NY, USA.
- Malcolm, R. L. and MacCarthy, P. (1992). Quantitative evaluation of XAD-8 and XAD-4 resins used in tandem for removing organic solutes from water. *Environment International*, 18(6):597–607.
- Malik, A. and Gleixner, G. (2013). Importance of microbial soil organic matter processing in dissolved organic carbon production. *FEMS Microbiology Ecology*, 86(1):139–148.
- Mangal, V., Stock, N. L., and Guéguen, C. (2016). Molecular characterization of phytoplankton dissolved organic matter (DOM) and sulfur components using high resolution Orbitrap mass spectrometry. *Analytical and Bioanalytical Chemistry*, 408(7):1891–1900.
- Mansell, R. L., Gross, G. G., Stöckigt, J., Franke, H., and Zenk, M. H. (1974). Purification and properties of cinnamyl alcohol dehydrogenase from higher plants involved in lignin biosynthesis. *Phytochemistry*, 13(11):2427–2435.

- Manzoni, S. and Katul, G. (2014). Invariant soil water potential at zero microbial respiration explained by hydrological discontinuity in dry soils. *Geophysical Research Letters*, 41(20):7151–7158.
- Manzoni, S. and Porporato, A. (2009). Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biology and Biochemistry*, 41(7):1355–1379.
- Mao, J., Cory, R. M., McKnight, D. M., and Schmidt-Rohr, K. (2007). Characterization of a nitrogen-rich fulvic acid and its precursor algae from solid state NMR. *Organic Geochemistry*, 38(8):1277–1292.
- Marschner, B. and Kalbitz, K. (2003). Controls of bioavailability and biodegradability of dissolved organic matter in soils. *Geoderma*, 113(3):211–235.
- Martens, D. A., Reedy, T. E., and Lewis, D. T. (2004). Soil organic carbon content and composition of 130-year crop, pasture and forest land-use managements. *Global Change Biology*, 10(1):65–78.
- Massicotte, P. (2017). **eemR**: Tools for Pre-Processing Emission-Excitation-Matrix (EEM) Fluorescence Data.
- Mazzilli, S. R., Kemanian, A. R., Ernst, O. R., Jackson, R. B., and Piñeiro, G. (2014). Priming of soil organic carbon decomposition induced by corn compared to soybean crops. *Soil Biology and Biochemistry*, 75:273–281.
- Mazzilli, S. R., Kemanian, A. R., Ernst, O. R., Jackson, R. B., and Piñeiro, G. (2015). Greater humification of belowground than aboveground biomass carbon into particulate soil organic matter in no-till corn and soybean crops. *Soil Biology and Biochemistry*, 85:22–30.
- McDowell, W., Zsolnay, Á., Aitkenhead-Peterson, J., Gregorich, E., Jones, D., Joemann, D., Kalbitz, K., Marschner, B., and Schwesig, D. (2006). A comparison of methods to determine the biodegradable dissolved organic matter (DOM) from different terrestrial sources. *Soil Biology and Biochemistry*, 38:1933–1942.
- McDowell, W. H. (2003). Dissolved organic matter in soils—future directions and unanswered questions. *Geoderma*, 113(3-4):179–186.
- McGill, W. B. and Cole, C. V. (1981). Comparative aspects of cycling of organic C, N, S and P through soil organic matter. *Geoderma*, 26(4):267–286.
- McKee, G. A., Soong, J. L., Caldéron, F., Borch, T., and Cotrufo, M. F. (2016). An integrated spectroscopic and wet chemical approach to investigate grass litter decomposition chemistry. *Biogeochemistry*, 128(1):107–123.

- McKnight, D. M., Boyer, E. W., Westerhoff, P. K., Doran, P. T., Kulbe, T., and Andersen, D. T. (2001). Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnology and Oceanography*, 46(1):38–48.
- McLaren, R. and Cameron, K. C. (1996). *Soil science: sustainable production and environmental protection*. Oxford University Press, 2<sup>nd</sup> edition.
- McLauchlan, K. K. and Hobbie, S. E. (2004). Comparison of Labile Soil Organic Matter Fractionation Techniques. *Soil Science Society of America Journal*, 68:1616–1625.
- McNally, S., Beare, M., Curtin, D., Tregurtha, C., Qiu, W., Kelliher, F., and Baldock, J. (2018). Assessing the vulnerability of organic matter to C mineralisation in pasture and cropping soils of New Zealand. *Soil Research*, 56(5):481–490.
- McNeill, J. R. and Winiwarter, V. (2004). Breaking the Sod: Humankind, History, and Soil. *Science*, 304(5677):1627–1629.
- Meadows, J. A. and Wargo, M. J. (2015). Carnitine in bacterial physiology and metabolism. *Microbiology (Reading, England)*, 161(6):1161–1174.
- Merckx, R., Dijkstra, A., den Hartog, A., and van Veen, J. A. (1987). Production of root-derived material and associated microbial growth in soil at different nutrient levels. *Biology and Fertility of Soils*, 5(2):126–132.
- Miano, T. M., Martin, J. P., and Sposito, G. (1988). Fluorescence Spectroscopy of Humic Substances. *Soil Science Society of America Journal*, 52:1016–1019.
- Miano, T. M. and Senesi, N. (1992). Synchronous excitation fluorescence spectroscopy applied to soil humic substances chemistry. *Science of The Total Environment*, 117-118:41–51.
- Michel, K., Matzner, E., and Dignac, M.-F. and Kögel-Knabner, I. (2006). Properties of dissolved organic matter related to soil organic matter quality and nitrogen additions in Norway spruce forest floors. *Geoderma*, 130(3-4):250–264.
- Miller, A. E., Schimel, J. P., Meixner, T., Sickman, J. O., and Melack, J. M. (2005). Episodic rewetting enhances carbon and nitrogen release from chaparral soils. *Soil Biology and Biochemistry*, 37(12):2195–2204.
- Milori, D. M. B. P., Galeti, H. V. A., Martin-Neto, L., Dieckow, J., González-Pérez, M., Bayer, C., and Salton, J. (2006). Organic matter study of whole soil samples using laser-induced fluorescence spectroscopy. *Soil Science Society of America Journal*, 70(1):57–63.

- Milori, D. M. B. P., Martin-Neto, L., Bayer, C., Mielniczuk, J., and Bagnato, V. S. (2002). Humification degree of soil humic acids determined by fluorescence spectroscopy. *Soil Science*, 167(11):739–749.
- Miltner, A., Kindler, R., Knicker, H., Richnow, H.-H., and Kästner, M. (2009). Fate of microbial biomass-derived amino acids in soil and their contribution to soil organic matter. *Organic Geochemistry*, 40(9):978–985.
- Minor, E. C., Swenson, M. M., Mattson, B. M., and Oyler, A. R. (2014). Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environmental Science: Processes & Impacts*, 16(9):2064–2079.
- Mobed, J. J., Hemmingsen, S. L., Autry, J. L., and McGown, L. B. (1996). Fluorescence characterization of IHSS humic substances: Total luminescence spectra with absorbance correction. *Environmental Science & Technology*, 30(10):3061–3065.
- Molloy, M. P., Donohoe, S., Brzezinski, E. E., Kilby, G. W., Stevenson, T. I., Baker, J. D., Goodlett, D. R., and Gage, D. A. (2005). Large-scale evaluation of quantitative reproducibility and proteome coverage using acid cleavable isotope coded affinity tag mass spectrometry for proteomic profiling. *Proteomics*, 5(5):1204–1208.
- Monod, J. (1947). The phenomenon of enzymatic adaptation and its bearings on problems of genetics and cellular differentiation. *Growth*, 11:223–289.
- Moore, T. and Dalva, M. (2001). Some controls on the release of dissolved organic carbon by plant tissues and soils. *Soil Science*, 166(1):38–47.
- Mooshammer, M., Wanek, W., Hämmerle, I., Fuchslueger, L., Hofhansl, F., Knoltsch, A., Schnecker, J., Takriti, M., Watzka, M., Wild, B., Keiblinger, K. M., Zechmeister-Boltenstern, S., and Richter, A. (2014). Adjustment of microbial nitrogen use efficiency to carbon:nitrogen imbalances regulates soil nitrogen cycling. *Nature Communications*, 5:3694.
- Morris, S. J. and Blackwood, C. B. (2007). 8 – The Ecology Of Soil Organisms. In Eldor, P. A., editor, *Soil Microbiology, Ecology and Biochemistry*, pages 195–229. Academic Press, San Diego, CA, USA, 3<sup>rd</sup> edition.
- Moyano, F. E., Manzoni, S., and Chenu, C. (2013). Responses of soil heterotrophic respiration to moisture availability: An exploration of processes and models. *Soil Biology and Biochemistry*, 59:72–85.



- Mueller, C. W., Gutsch, M., Kothieringer, K., Leifeld, J., Rethemeyer, J., Brueggemann, N., and Kögel-Knabner, I. (2014). Bioavailability and isotopic composition of CO<sub>2</sub> released from incubated soil organic matter fractions. *Soil Biology and Biochemistry*, 69:168–178.
- Müller, M., Alewell, C., and Hagedorn, F. (2009). Effective retention of litter-derived dissolved organic carbon in organic layers. *Soil Biology and Biochemistry*, 41(6):1066–1074.
- Murphy, D. V., Macdonald, A. J., Stockdale, E. A., Goulding, K. W. T., Fortune, S., Gaunt, J. L., Poulton, P. R., Wakefield, J. A., Webster, C. P., and Wilmer, W. S. (2000). Soluble organic nitrogen in agricultural soils. *Biology and Fertility of Soils*, 30(5):374–387.
- Murphy, K. R., Stedmon, C. A., Graeber, D., and Bro, R. (2013). Fluorescence spectroscopy and multi-way techniques. PARAFAC. *Analytical Methods*, 5(23):6557–6566.
- Murray, K. K., Boyd, R. K., Eberlin, M. N., Langley, J. G., Li, L., and Naito, Y. (2013). Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure and Applied Chemistry*, 85(7):1515–1609.
- Nannipieri, P. and Eldor, P. A. (2009). The chemical and functional characterization of soil N and its biotic components. *Soil Biology and Biochemistry*, 41(12):2357–2369.
- Neff, J. C. and Asner, G. P. (2001). Dissolved Organic Carbon in Terrestrial Ecosystems: Synthesis and a Model. *Ecosystems*, 4(1):29–48.
- Newman, D. K. and Kolter, R. (2000). A role for excreted quinones in extracellular electron transfer. *Nature*, 405:94.
- Ni, J. and Pignatello, J. J. (2018). Charge-assisted hydrogen bonding as a cohesive force in soil organic matter: water solubility enhancement by addition of simple carboxylic acids. *Environmental Science: Processes & Impacts*, 20(9):1225–1233.
- Nierop, K. G. J., Pulleman, M. M., and Marinissen, J. C. Y. (2001). Management induced organic matter differentiation in grassland and arable soil: a study using pyrolysis techniques. *Soil Biology and Biochemistry*, 33(6):755–764.
- Novak, J. M., Mills, G. L., and Bertsch, P. M. (1992). Estimating the Percent Aromatic Carbon in Soil and Aquatic Humic Substances Using Ultraviolet Absorbance Spectrometry. *Journal of Environmental Quality*, 21:144–147.

- Ohno, T. (2002). Fluorescence Inner-Filtering Correction for Determining the Humification Index of Dissolved Organic Matter. *Environmental Science & Technology*, 36(4):742–746.
- Ohno, T. and Bro, R. (2006). Dissolved Organic Matter Characterization Using Multiway Spectral Decomposition of Fluorescence Landscapes. *Soil Science Society of America Journal*, 70:2028–2037.
- Ohno, T. and First, P. R. (1998). Assessment of the Folin and Ciocalteu's Method for Determining Soil Phenolic Carbon. *Journal of Environmental Quality*, 27:776–782.
- Ohno, T., He, Z., Sleighter, R. L., Honeycutt, C. W., and Hatcher, P. G. (2010). Ultrahigh Resolution Mass Spectrometry and Indicator Species Analysis to Identify Marker Components of Soil- and Plant Biomass-Derived Organic Matter Fractions. *Environmental Science & Technology*, 44(22):8594–8600.
- Ohno, T., He, Z., Tazisong, I. A., and Senwo, Z. N. (2009). Influence of Tillage, Cropping, and Nitrogen Source on the Chemical Characteristics of Humic Acid, Fulvic Acid, and Water-Soluble Soil Organic Matter Fractions of a Long-Term Cropping System Study. *Soil Science*, 174(12):652–660.
- Pan, H., Yu, H., Song, Y., Zhu, L., Liu, R., and Du, E. (2017). Tracking fluorescent components of dissolved organic matter from soils in large-scale irrigated area. *Environmental Science and Pollution Research*, 24(7):6563–6571.
- Parlanti, E., Wörz, K., Geoffroy, L., and Lamotte, M. (2000). Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Organic Geochemistry*, 31(12):1765–1781.
- Parton, W. J., Grosso, S. D., Plante, A. F., and Adair, E. C. (2015). Modeling the Dynamics of Soil Organic Matter and Nutrient Cycling. In Paul, E. A., editor, *Soil Microbiology, Ecology and Biochemistry*, pages 505–537. Academic Press, Amsterdam, 4<sup>th</sup> edition.
- Paul, E. A. (2015). *Soil Microbiology, Ecology and Biochemistry*. Academic Press, Amsterdam, 4<sup>th</sup> edition.
- Paustian, K., Six, J., Elliott, E. T., and Hunt, H. W. (2000). Management options for reducing CO<sub>2</sub> emissions from agricultural soils. *Biogeochemistry*, 48(1):147–163.

- Perdue, E. M. (1984). Analytical constraints on the structural features of humic substances. *Geochimica et Cosmochimica Acta*, 48(7):1435–1442.
- Perdue, E. M. and Ritchie, J. D. (2003). Dissolved organic matter in freshwaters. In Turekian, H. D. H. K., editor, *Treatise on Geochemistry*, pages 273–318. Pergamon, Oxford, UK.
- Perry, R. H., Cooks, R. G., and Noll, R. J. (2008). Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrometry Reviews*, 27(6):661–699.
- Peterson, S. and Jeffers, R. W. (1952). Equilibria between Aliphatic Acids and a Strong Base Anion Exchanger. *Journal of the American Chemical Society*, 74(6):1605–1606.
- Peuravuori, J., Koivikko, R., and Pihlaja, K. (2002). Characterization, differentiation and classification of aquatic humic matter separated with different sorbents: synchronous scanning fluorescence spectroscopy. *Water Research*, 36(18):4552–4562.
- Pisani, O., Hills, K., Courtier-Murias, D., Haddix, M., Paul, E., T. Conant, R., J. Simpson, A., Arhonditsis, G., and Simpson, M. (2014). Accumulation of aliphatic compounds in soil with increasing mean annual temperature. *Organic Geochemistry*, 76:118–127.
- Pluskal, T., Uehara, T., and Yanagida, M. (2012). Highly Accurate Chemical Formula Prediction Tool Utilizing High-Resolution Mass Spectra, MS/MS Fragmentation, Heuristic Rules, and Isotope Pattern Matching. *Analytical Chemistry*, 84(10):4396–4403.
- Poirer, N., Derenne, S., Balesdent, J., Chenu, C., Bardoux, G., Mariotti, A., and Largeau, C. (2006). Dynamics and origin of the non-hydrolysable organic fraction in a forest and a cultivated temperate soil, as determined by isotopic and microscopic studies. *European Journal of Soil Science*, 57(5):719–730.
- Poirier, N., Sohi, S. P., Gaunt, J. L., Mahieu, N., Randall, E. W., Powlson, D. S., and Evershed, R. P. (2005). The chemical composition of measurable soil organic matter pools. *Organic Geochemistry*, 36(8):1174–1189.
- Pörtl, K., Zechmeister-Boltenstern, S., Wanek, W., Ambus, P., and Berger, T. W. (2007). Natural  $^{15}\text{N}$  abundance of soil N pools and  $\text{N}_2\text{O}$  reflect the nitrogen dynamics of forest soils. *Plant and Soil*, 295(1):79–94.

- Powlson, D. S. and Jenkinson, D. S. (1976). The effects of biocidal treatments on metabolism in soil—II. Gamma irradiation, autoclaving, air-drying and fumigation. *Soil Biology and Biochemistry*, 8(3):179–188.
- Qin, Y., Zhu, H., Zhang, M., Zhang, H., Xiang, C., and Li, B. (2016). GC-MS analysis of membrane-graded fulvic acid and its activity on promoting wheat seed germination. *Molecules*, 21:1363.
- Qualls, R. and Haines, B. L. (1992). Biodegradability of Dissolved Organic Matter in Forest Throughfall, Soil Solution, and Stream Water. *Soil Science Society of America Journal - SSSAJ*, 56:586–591.
- Qualls, R. G. (2005). Biodegradability of Fractions of Dissolved Organic Carbon Leached from Decomposing Leaf Litter. *Environmental Science & Technology*, 39(6):1616–1622.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raphael, J. P. A., Calonego, J. C., Milori, D. M. B. P., and Rosolem, C. A. (2016). Soil organic matter in crop rotations under no-till. *Soil and Tillage Research*, 155:45–53.
- Redmile-Gordon, M. A., Evershed, R. P., Hirsch, P. R., White, R. P., and Goulding, K. W. T. (2015). Soil organic matter and the extracellular microbial matrix show contrasting responses to C and N availability. *Soil Biology and Biochemistry*, 88:257–267.
- Reid, I. D. and Deschamps, A. M. (1991). Nutritional regulation of synthetic lignin (DHP) degradation by the selective white-rot fungus *Phlebia (Merulius) tremellosa*: effects of glucose and other cosubstrates. *Canadian Journal of Botany*, 69(1):147–155.
- Reutemann, P. and Holmes, G. (2015). Big Data with ADAMS. In Fan, W., Bifet, A., Yang, Q., and Yu, P. S., editors, *BIGMINE’15 Proceedings of the 4<sup>th</sup> International Workshop on Big Data Streams and Heterogeneous Source Mining: Algorithms Systems, Programming Models and Applications*, volume 41, pages 5–8. JMLR.org.
- Reutemann, P. and Vanschoren, J. (2012). Scientific Workflow Management with ADAMS. In Flach, P. A., De Bie, T., and Cristianini, N., editors, *Machine Learning and Knowledge Discovery in Databases (ECML-PKDD), Part II*, volume 7524, pages 833–837. Springer, Berlin, Heidelberg.

- Rinot, O., Osterholz, W. R., Castellano, M. J., Linker, R., Liebman, M., and Shaviv, A. (2018). Excitation-Emission-Matrix Fluorescence Spectroscopy of Soil Water Extracts to Predict Nitrogen Mineralization Rates. *Soil Science Society of America Journal*, 82:126–135.
- Ríos, H. E. and Urzúa, M. (2001). Ion Exchange Equilibrium in Polyelectrolyte Solutions through the  $\text{CHCl}_3$ /Water Interface. *Journal of Colloid and Interface Science*, 242(2):460–464.
- Rivas-Ubach, A., Liu, Y., Bianchi, T. S., Tolić, N., Jansson, C., and Paša-Tolić, L. (2018). Moving beyond the van Krevelen Diagram: A New Stoichiometric Approach for Compound Classification in Organisms. *Analytical Chemistry*, 90(10):6152–6160.
- Roberts, P., Stockdale, R., Khalid, M., Iqbal, Z., and Jones, D. L. (2009). Carbon-to-nitrogen ratio is a poor predictor of low molecular weight organic nitrogen mineralization in soil. *Soil Biology and Biochemistry*, 41(8):1750–1752.
- Rodrigo, A., Recous, S., Neel, C., and Mary, B. (1997). Modelling temperature and moisture effects on C–N transformations in soils: comparison of nine models. *Ecological Modelling*, 102:325–339.
- Rodríguez, F. J. and Núñez, L. A. (2011). Characterization of aquatic humic substances. *Water and Environment Journal*, 25(2):163–170.
- Rodríguez, F. J., Schlenger, P., and García-Valverde, M. (2014). A comprehensive structural evaluation of humic substances using several fluorescence techniques before and after ozonation. Part I: Structural characterization of humic substances. *Science of The Total Environment*, 476-477:718–730.
- Ros, G. H., Hanegraaf, M. C., Hoffland, E., and van Riemsdijk, W. H. (2011a). Predicting soil N mineralization: Relevance of organic matter fractions and soil properties. *Soil Biology and Biochemistry*, 43(8):1714–1722.
- Ros, G. H., Hoffland, E., van Kessel, C., and Temminghoff, E. J. M. (2009). Extractable and dissolved soil organic nitrogen – A quantitative assessment. *Soil Biology and Biochemistry*, 41(6):1029–1039.
- Ros, G. H., Temminghoff, E. J. M., and Hoffland, E. (2011b). Nitrogen mineralization: a review and meta-analysis of the predictive value of soil tests. *European Journal of Soil Science*, 62(1):162–173.

- Rosa, A. H., Simões, M. L., Camargo de Oliveira, L., Rocha, J. C., Neto, L. M., and Milori, D. M. B. P. (2005). Multimethod study of the degree of humification of humic substances extracted from different tropical soil profiles in Brazil's Amazonian region. *Geoderma*, 127(1):1–10.
- Roth, S., Jung, K., Jung, H., Hommel, R. K., and Kleber, H.-P. (1994). Crotonobetaine reductase from *Escherichia coli* – a new inducible enzyme of anaerobic metabolism of L(-)-carnitine. *Antonie van Leeuwenhoek*, 65(1):63–69.
- Roth, V.-N., Dittmar, T., Gaupp, R., and Gleixner, G. (2013). Latitude and pH driven trends in the molecular composition of DOM across a north south transect along the Yenisei River. *Geochimica et Cosmochimica Acta*, 123:93–105.
- Roth, V.-N., Dittmar, T., Gaupp, R., and Gleixner, G. (2015). The molecular composition of dissolved organic matter in forest soils as a function of pH and temperature. *PLoS ONE*, 10(3):e0119188.
- Rouwane, A., Grybos, M., Bourven, I., Rabiet, M., and Guibaud, G. (2018). Waterlogging and soil reduction affect the amount and apparent molecular weight distribution of dissolved organic matter in wetland soil: a laboratory study. *Soil Research*, 56(1):28–38.
- Rumpel, C., Eusterhues, K., and Kögel-Knabner, I. (2004). Location and chemical composition of stabilized organic carbon in topsoil and subsoil horizons of two acid forest soils. *Soil Biology and Biochemistry*, 36(1):177–190.
- Saadi, I., Borisover, M., Armon, R., and Laor, Y. (2006). Monitoring of effluent DOM biodegradation using fluorescence, UV and DOC measurements. *Chemosphere*, 63(3):530–539.
- Salkind, N. J. (2007). *Encyclopedia of measurement and statistics*, volume 1. Sage.
- Sanderman, J., Maddern, T., and Baldock, J. (2014). Similar composition but differential stability of mineral retained organic matter across four classes of clay minerals. *Biogeochemistry*, 121(2):409–424.
- Schimel, J., Balser, T. C., and Wallenstein, M. (2007). Microbial stress-response physiology and its implications for ecosystem function. *Ecology*, 88(6):1386–1394.
- Schmidt, F., Elvert, M., Koch, B. P., Witt, M., and Hinrichs, K.-U. (2009). Molecular characterization of dissolved organic matter in pore water of continental shelf sediments. *Geochimica et Cosmochimica Acta*, 73(11):3337–3358.

- Schmidt, M. W., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D. A. C., Nannipieri, P., Rasse, D. P., Weiner, S., and Trumbore, S. E. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, 478:49–56.
- Schnabel, R. R., Dell, C. J., and Shaffer, J. A. (2002). Filter, inoculum and time effects on measurements of biodegradable water soluble organic carbon in soil. *Soil Biology and Biochemistry*, 34(5):737–739.
- Schneckenburger, T., Lattao, C., Pignatello, J. J., Schaumann, G. E., Thiele-Bruhn, S., Cao, X., and Mao, J. (2012). Preparation and characterization of humic acid cross-linked with organic bridging groups. *Organic Geochemistry*, 47:132–138.
- Scow, K. M., Schmidt, S. K., and Alexander, M. (1989). Kinetics of biodegradation of mixtures of substrates in soil. *Soil Biology and Biochemistry*, 21(5):703–708.
- Seasholtz, M. B. and Kowalski, B. (1993). The parsimony principle applied to multivariate calibration. *Analytica Chimica Acta*, 277(2):165–177.
- Segnini, A., Carvalho, J. L. N., Bolonhezi, D., Milori, D. M. B. P., da Silva, W. T. L., Simões, M. L., Cantarella, H., de Maria, I. C., and Martin-Neto, L. (2013). Carbon stock and humification index of organic matter affected by sugarcane straw and soil management.
- Seidel, M., Beck, M., Riedel, T., Waska, H., Suryaputra, I. G. N. A., Schnetger, B., Niggemann, J., Simon, M., and Dittmar, T. (2014). Biogeochemistry of dissolved organic matter in an anoxic intertidal creek bank. *Geochimica et Cosmochimica Acta*, 140:418–434.
- Seifert, A.-G., Roth, V.-N., Dittmar, T., Gleixner, G., Breuer, L., Houska, T., and Marxsen, J. (2016). Comparing molecular composition of dissolved organic matter in soil and stream water: Influence of land use and chemical characteristics. *Science of The Total Environment*, 571:142–152.
- Semple, K. T., Doick, K. J., Jones, K. C., Burauel, P., Craven, A., and Harms, H. (2004). Defining bioavailability and bioaccessibility of contaminated soil and sediment is complicated. *Environmental Science & Technology*, 38(12):228A–231A.
- Senesi, N. (1990). Molecular and quantitative aspects of the chemistry of fulvic acid and its interactions with metal ions and organic chemicals: Part II. The fluorescence spectroscopy approach. *Analytica Chimica Acta*, 232:77–106.

- Senesi, N. and D'Orazio, V. (2005). Fluorescence spectroscopy. In Hillel, D., editor, *Encyclopedia of Soils in the Environment*, pages 35–52. Elsevier Science, London, UK.
- Senesi, N., Miano, T. M., Provenzano, M. R., and Brunetti, G. (1991). Characterization, differentiation, and classification of humic substances by fluorescence spectroscopy. *Soil Science*, 152(4):259–271.
- Serventi, F., Ramazzina, I., Lamberto, I., Puggioni, V., Gatti, R., and Percudani, R. (2010). Chemical Basis of Nitrogen Recovery through the Ureide Pathway: Formation and Hydrolysis of *S*-Ureidoglycine in Plants and Bacteria. *ACS Chemical Biology*, 5(2):203–214.
- Shahbaz, M., Kuzyakov, Y., and Heitkamp, F. (2017). Decrease of soil organic matter stabilization with increasing inputs: mechanisms and controls. *Geoderma*, 304:76–82.
- Shahzad, T., Chenu, C., Repinçay, C., Mougin, C., Ollier, J.-L., and Fontaine, S. (2012). Plant clipping decelerates the mineralization of recalcitrant soil organic matter under multiple grassland species. *Soil Biology and Biochemistry*, 51:73–80.
- Shalizi, C. R. (2017). *Advanced Data Analysis from an Elementary Point of View*. Cambridge University Press.
- Shiau, Y.-J., Wang, H.-C., Chen, T.-H., Jien, S.-H., Tian, G., and Chiu, C.-Y. (2017). Improvement in the biochemical and chemical properties of badland soils by thorny bamboo. *Scientific Reports*, 7:40561.
- Shin, I., Percudani, R., and Rhee, S. (2012). Structural and Functional Insights into (*S*)-Ureidoglycine Aminohydrolase, Key Enzyme of Purine Catabolism in *Arabidopsis thaliana*. *The Journal of Biological Chemistry*, 287(22):18796–18805.
- Shirato, Y., Hakamata, T., and Taniyama, I. (2004). Modified Rothamsted carbon model for Andosols and its validation: changing humus decomposition rate constant with pyrophosphate-extractable Al. *Soil Science and Plant Nutrition*, 50(1):149–158.
- Simon, C., Roth, V.-N., Dittmar, T., and Gleixner, G. (2018). Molecular Signals of Heterogeneous Terrestrial Environments Identified in Dissolved Organic Matter: A Comparative Analysis of Orbitrap and Ion Cyclotron Resonance Mass Spectrometers. *Frontiers in Earth Science*, 6:138.
- Singer, P. C. (1999). Humic substances as precursors for potentially harmful disinfection by-products. *Water Science and Technology*, 40(9):25–30.



- Sinsabaugh, R. L. and Follstad Shah, J. J. (2012). Ecoenzymatic Stoichiometry and Ecological Theory. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):313–343.
- Sisti, C. P. J., dos Santos, H. P., Kohhann, R., Alves, B. J. R., Urquiaga, S., and Boddey, R. M. (2004). Change in carbon and nitrogen stocks in soil under 13 years of conventional or zero tillage in southern Brazil. *Soil and Tillage Research*, 76(1):39–58.
- Six, J., Bossuyt, H., Degryze, S., and Denef, K. (2004). A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. *Soil and Tillage Research*, 79(1):7–31.
- Six, J., Elliott, E. T., and Paustian, K. (1999). Aggregate and Soil Organic Matter Dynamics under Conventional and No-Tillage Systems. *Soil Science Society of America Journal*, 63:1350–1358.
- Sleighter, R. L. and Hatcher, P. G. (2008). Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Marine Chemistry*, 110(3):140–152.
- Sleighter, R. L., Liu, Z., Xue, J., and Hatcher, P. G. (2010). Multivariate Statistical Approaches for the Characterization of Dissolved Organic Matter Analyzed by Ultrahigh Resolution Mass Spectrometry. *Environmental Science & Technology*, 44(19):7576–7582.
- Sleno, L., Volmer, D. A., and Marshall, A. G. (2005). Assigning product ions from complex MS/MS spectra: The importance of mass uncertainty and resolving power. *Journal of the American Society for Mass Spectrometry*, 16(2):183–198.
- Smucker, A. J. M., Park, E.-J., Dorner, J., and Horn, R. (2007). Soil micropore development and contributions to soluble carbon transport within macroaggregates. *Vadose Zone Journal*, 6(2):282–290.
- Sokol, N. W. and Bradford, M. A. (2019). Microbial formation of stable soil carbon is more efficient from belowground than aboveground input. *Nature Geoscience*, 12(1):46–53.
- Sollins, P., Homann, P., and Caldwell, B. A. (1996). Stabilization and destabilization of soil organic matter: mechanisms and controls. *Geoderma*, 74(1):65–105.
- Sollins, P., Swanston, C., Kleber, M., Filley, T., Kramer, M., Crow, S., Caldwell, B. A., Lajtha, K., and Bowden, R. (2006). Organic C and N stabilization in

- a forest soil: evidence from sequential density fractionation. *Soil Biology and Biochemistry*, 38(11):3313–3324.
- Sollins, P., Swanston, C., and Kramer, M. (2007). Stabilization and destabilization of soil organic matter—a new focus. *Biogeochemistry*, 85(1):1–7.
- Sparling, G., Vojvodić-Vuković, M., and Schipper, L. A. (1998). Hot-water-soluble C as a simple measure of labile soil organic matter: The relationship with microbial biomass C. *Soil Biology and Biochemistry*, 30(10):1469–1472.
- Stapleton, R. D., Savage, D. C., Sayler, G. S., and Stacey, G. (1998). Biodegradation of aromatic hydrocarbons in an extremely acidic environment. *Applied and Environmental Microbiology*, 64(11):4180–4184.
- Stedmon, C. A. and Bro, R. (2008). Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnology and Oceanography: Methods*, 6(11):572–579.
- Stedmon, C. A., Markager, S., and Bro, R. (2003). Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry*, 82(3):239–254.
- Steenwerth, K. L., Jackson, L. E., Calderón, F. J., Scow, K. M., and Rolston, D. E. (2005). Response of microbial community composition and activity in agricultural and grassland soils after a simulated rainfall. *Soil Biology and Biochemistry*, 37(12):2249–2262.
- Stenson, A. C., Marshall, A. G., and Cooper, W. T. (2003). Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Analytical Chemistry*, 75(6):1275–1284.
- Stevenson, F. J. (1994). *Humus chemistry: genesis, composition, reactions*. John Wiley and Sons, New York.
- Stres, B. and Tiedje, J. M. (2006). New Frontiers in Soil Microbiology: How To Link Structure and Function of Microbial Communities? In Nannipieri, P. and Smalla, K., editors, *Nucleic Acids and Proteins in Soil*, pages 1–22. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Sugai, S. F. and Schimel, J. P. (1993). Decomposition and biomass incorporation of  $^{14}\text{C}$ -labeled glucose and phenolics in taiga forest floor: effect of substrate quality, successional state, and season. *Soil Biology and Biochemistry*, 25(10):1379–1389.

- Swenson, T. L., Karaoz, U., Swenson, J. M., Bowen, B. P., and Northen, T. R. (2018). Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. *Nature Communications*, 9(1):19.
- Swift, R. S. and Posner, A. M. (1972). Autoxidation of humic acid under alkaline conditions. *Journal of Soil Science*, 23(4):381–393.
- Takada, M., Rabemanolontsoa, H., Minami, E., and Saka, S. (2018). Characterization of lignin-derived products from various lignocellulosics as treated by semi-flow hot-compressed water. *Journal of Wood Science*, 64(6):802–809.
- Talibart, R., Jebbar, M., Gouesbet, G., Himdi-Kabbab, S., Wróblewski, H., Blanco, C., and Bernard, T. (1994). Osmoadaptation in rhizobia: ectoine-induced salt tolerance. *Journal of Bacteriology*, 176(17):5210–5217.
- Tamamushi, B. and Tamaki, K. (1959). Adsorption of long-chain electrolytes at the solid/liquid interface. Part 3.—The adsorption on ion-exchange resins. *Transactions of the Faraday Society*, 55(0):1013–1016.
- Tanford, C. (1980). *The hydrophobic effect: formation of micelles and biological membranes*. John Wiley & Sons, New York, USA, 2<sup>nd</sup> edition.
- Teramoto, H., Tanaka, H., and Wariishi, H. (2004). Degradation of 4-nitrophenol by the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *Applied Microbiology and Biotechnology*, 66(3):312–317.
- Tfaily, M. M., Corbett, J. E., Wilson, R., Chanton, J. P., Glaser, P. H., Cawley, K. M., Jaffé, R., and Cooper, W. T. (2015). Utilization of PARAFAC-Modeled Excitation-Emission Matrix (EEM) Fluorescence Spectroscopy to Identify Biogeochemical Processing of Dissolved Organic Matter in a Northern Peatland. *Photochemistry and Photobiology*, 91:684–695.
- Tfaily, M. M., Hamdan, R., Corbett, J. E., Chanton, J. P., Glaser, P. H., and Cooper, W. T. (2013). Investigating dissolved organic matter decomposition in northern peatlands using complimentary analytical techniques. *Geochimica et Cosmochimica Acta*, 112:116–129.
- Thomas, G. W. (1982). Exchangeable cations. In Page, A. L., Miller, R. H., and Keeney, D. R., editors, *Methods of soil analysis. Part 2*, pages 159–164. American Society of Agronomy, Madison, WI, USA.
- Thurman, E. M. (1985). *Organic geochemistry of natural waters*. Martinus Nijhoff/Dr. W. Junk Publishers, Dordrecht, The Netherlands.

- Thygesen, L. G., Rinnan, Å., Barsberg, S., and Møller, J. K. S. (2004). Stabilizing the PARAFAC decomposition of fluorescence spectra by insertion of zeros outside the data area. *Chemometrics and Intelligent Laboratory Systems*, 71(2):97–106.
- Tian, J., Fan, M., Guo, J., Marschner, P., Li, X., and Kuzyakov, Y. (2012). Effects of land use intensity on dissolved organic carbon properties and microbial community structure. *European Journal of Soil Biology*, 52:67–72.
- Tisdall, J. M., Cockroft, B., and Uren, N. C. (1978). The stability of soil aggregates as affected by organic materials, microbial activity and physical disruption. *Soil Research*, 16(1):9–17.
- Traina, S. J., Novak, J., and Smeck, N. E. (1990). An Ultraviolet Absorbance Method of Estimating the Percent Aromatic Carbon Content of Humic Acids. *Journal of Environmental Quality*, 19:151–153.
- Tsai, C. S. (1967). Spontaneous decarboxylation of oxalacetic acid. *Canadian Journal of Chemistry*, 45(8):873–880.
- Uchida, Y., Nishimura, S., and Akiyama, H. (2012). The relationship of water-soluble carbon and hot-water-soluble carbon with soil respiration in agricultural fields. *Agriculture, Ecosystems & Environment*, 156:116–122.
- Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., and Markley, J. L. (2008). BioMagRes-Bank. *Nucleic Acids Research*, 36(Database issue):D402–D408.
- van Genuchten, M. (1980). A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Science Society of America Journal*, 44:892–898.
- van Genuchten, M. T., Leij, F. J., and Yates, S. R. (1991). The RETC code for quantifying the hydraulic functions of unsaturated soils. *USEPA Report 600/2-91/065*.
- Van Gestel, M., Merckx, R., and Vlassak, K. (1993a). Microbial biomass and activity in soils with fluctuating water content. *Geoderma*, 56:617–626.
- Van Gestel, M., Merckx, R., and Vlassak, K. (1993b). Soil drying and rewetting and the turnover of  $^{14}\text{C}$ -labelled plant residues: First order decay rates of biomass and non-biomass  $^{14}\text{C}$ . *Soil Biology and Biochemistry*, 25(1):125–134.

- van Kessel, C., Clough, T., and van Groenigen, J. W. (2009). Dissolved Organic Nitrogen: An Overlooked Pathway of Nitrogen Loss from Agricultural Systems? *Journal of Environmental Quality*, 38:393–401.
- Vergnoux, A., Di Rocco, R., Domeizel, M., Guiliano, M., Doumenq, P., and Théraulaz, F. (2011). Effects of forest fires on water extractable organic matter and humic substances from Mediterranean soils: UV-vis and fluorescence spectroscopy approaches. *Geoderma*, 160(3):434–443.
- Verkh, Y., Rozman, M., and Petrovic, M. (2018). A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment. *Chemosphere*, 200:397–404.
- Vogelmann, E. S., Reichert, J. M., Prevedello, J., Awe, G. O., and Mataix-Solera, J. (2013). Can occurrence of soil hydrophobicity promote the increase of aggregates stability? *CATENA*, 110:24–31.
- Volk, C. J., Volk, C. B., and Kaplan, L. A. (1997). Chemical composition of biodegradable dissolved organic matter in streamwater. *Limnology and Oceanography*, 42(1):39–44.
- von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E., and Marschner, B. (2007). SOM fractionation methods: Relevance to functional pools and to stabilization mechanisms. *Soil Biology and Biochemistry*, 39(9):2183–2207.
- von Lützow, M., Leifeld, J., Kainz, M., Kögel-Knabner, I., and Munch, J. C. (2002). Indications for soil organic matter quality in soils under different management. *Geoderma*, 105(3):243–258.
- Wachinski, A. M. (2016). *Environmental Ion Exchange: Principles and Design*. CRC Press, Boca Raton, FL, USA, 2<sup>nd</sup> edition.
- Waksman, S. A. (1932). *Humus: Origin, Chemical Composition, and Importance in Nature*. Williams & Wilkins, Baltimore.
- Walworth, J. L. (1992). Soil Drying and Rewetting, or Freezing and Thawing, Affects Soil Solution Composition. *Soil Science Society of America Journal*, 56:433–437.
- Wander, M. (2004). Soil organic matter fractions and their relevance to soil function. In Magdoff, F. and Weil, R. R., editors, *Soil Organic Matter in Sustainable Agriculture*, pages 67–102. CRC Press, Boca Raton, FL, USA.

- Wang, X., Xia, K., Yang, X., and Tang, C. (2019). Growth strategy of microbes on mixed carbon sources. *Nature Communications*, 10(1):1279.
- Wardle, D. A., Bonner, K. I., Barker, G. M., Yeates, G. W., Nicholson, K. S., Bardgett, R. D., Watson, R. N., and Ghani, A. (1999). Plant removals in perennial grassland: vegetation dynamics, decomposers, soil biodiversity, and ecosystem properties. *Ecological Monographs*, 69(4):535–568.
- Warren, C. R. (2013a). High diversity of small organic N observed in soil water. *Soil Biology and Biochemistry*, 57:444–450.
- Warren, C. R. (2013b). Quaternary ammonium compounds can be abundant in some soils and are taken up as intact molecules by plants. *New Phytologist*, 198(2):476–485.
- Warren, C. R. (2014). Response of osmolytes in soil to drying and rewetting. *Soil Biology and Biochemistry*, 70:22–32.
- Warren, C. R. (2019). Isotope pool dilution reveals rapid turnover of small quaternary ammonium compounds. *Soil Biology and Biochemistry*, 131:90–99.
- Webster, R. (2001). Statistics to support soil research and their presentation. *European Journal of Soil Science*, 52(2):331–340.
- Webster, R. (2008). Soil Sampling and Methods of Analysis. *European Journal of Soil Science*, 59(5):1010–1011.
- Weishaar, J. L., Aiken, G. R., Bergamaschi, B. A., Fram, M. S., Fujii, R., and Mopper, K. (2003). Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon. *Environmental Science & Technology*, 37(20):4702–4708.
- Wells, M. J. and Stretz, H. A. (2019). Supramolecular architectures of natural organic matter. *Science of The Total Environment*, 671:1125–1133.
- Wells, M. J. M., Mullins, G. A., Bell, K. Y., Da Silva, A. K., and Navarrete, E. M. (2017). Fluorescence and Quenching Assessment (EEM-PARAFAC) of de Facto Potable Reuse in the Neuse River, North Carolina, United States. *Environmental Science & Technology*, 51(23):13592–13602.
- Werth, M. and Kuzyakov, Y. (2010).  $^{13}\text{C}$  fractionation at the root-microorganisms-soil interface: a review and outlook for partitioning studies. *Soil Biology and Biochemistry*, 42(9):1372–1384.

- West, A. W., Sparling, G. P., Feltham, C. W., and Reynolds, J. (1992). Microbial activity and survival in soils dried at different rates. *Soil Research*, 30(2):209–222.
- Whitehead, D. C. (1964). Identification of *p*-Hydroxybenzoic, Vanillic, *p*-Coumaric and Ferulic Acids in Soils. *Nature*, 202(4930):417–418.
- Wilkinson, A., Hill, P. W., Farrar, J. F., Jones, D. L., and Bardgett, R. D. (2014). Rapid microbial uptake and mineralization of amino acids and peptides along a grassland productivity gradient. *Soil Biology and Biochemistry*, 72:75–83.
- Wilson, H. F. and Xenopoulos, M. A. (2009). Effects of agricultural land use on the composition of fluvial dissolved organic matter. *Nature Geoscience*, 2:37–41.
- Wilson, I. D., Adlard, E. R., Cooke, M., and Poole, C. F. (2000). *Encyclopedia of Separation Science*. Academic Press, Boca Raton, FL, USA, 1<sup>st</sup> edition.
- Wortel, M. T., Peters, H., Hulshof, J., Teusink, B., and Bruggeman, F. J. (2014). Metabolic states with maximal specific rate carry flux through an elementary flux mode. *The FEBS Journal*, 281(6):1547–1555.
- Wu, J. and Brookes, P. C. (2005). The proportional mineralisation of microbial biomass and organic matter caused by air-drying and rewetting of a grassland soil. *Soil Biology and Biochemistry*, 37(3):507–515.
- Xiao, H., Krauss, M., Floehr, T., Yan, Y., Bahlmann, A., Eichbaum, K., Brinkmann, M., Zhang, X., Yuan, X., Brack, W., and Hollert, H. (2016). Effect-Directed Analysis of Aryl Hydrocarbon Receptor Agonists in Sediments from the Three Gorges Reservoir, China. *Environmental Science & Technology*, 50(20):11319–11328.
- Zang, X., van Heemst, J. D., Dria, K. J., and Hatcher, P. G. (2000). Encapsulation of protein in humic acid from a histosol as an explanation for the occurrence of organic nitrogen in soil and sediment. *Organic Geochemistry*, 31(7-8):679–695.
- Zanghellini, J., Ruckerbauer, D. E., Hanscho, M., and Jungreuthmayer, C. (2013). Elementary flux modes in a nutshell: Properties, calculation and applications. *Biotechnology Journal*, 8(9):1009–1016.
- Zech, W., Senesi, N., Guggenberger, G., Kaiser, K., Lehmann, J., Miano, T. M., Miltner, A., and Schroth, G. (1997). Factors controlling humification and mineralization of soil organic matter in the tropics. *Geoderma*, 79(1):117–161.
- Zhang, C. and Bennett, G. N. (2005). Biodegradation of xenobiotics by anaerobic bacteria. *Applied Microbiology and Biotechnology*, 67(5):600–618.

- Zhang, Y. and Ezeji, T. C. (2014). Elucidating and alleviating impacts of lignocellulose-derived microbial inhibitors on *Clostridium beijerinckii* during fermentation of *Miscanthus giganteus* to butanol. *Journal of Industrial Microbiology & Biotechnology*, 41(10):1505–1516.
- Zornoza, R., Guerrero, C., Mataix-Solera, J., Arcenegui, V., García-Orenes, F., and Mataix-Beneyto, J. (2007). Assessing the effects of air-drying and rewetting pre-treatment on soil microbial biomass, basal respiration, metabolic quotient and soluble carbon under Mediterranean conditions. *European Journal of Soil Biology*, 43(2):120–129.
- Zou, C.-S., Mo, M.-H., Gu, Y.-Q., Zhou, J.-P., and Zhang, K.-Q. (2007). Possible contributions of volatile-producing bacteria to soil fungistasis. *Soil Biology and Biochemistry*, 39(9):2371–2379.
- Zsolnay, Á. (1996). Chapter 4 – Dissolved Humus in Soil Waters. In Piccolo, A., editor, *Humic Substances in Terrestrial Ecosystems*, pages 171–223. Elsevier Science B.V., Amsterdam.
- Zsolnay, Á. (2003). Dissolved organic matter: artefacts, definitions, and functions. *Geoderma*, 113(3):187–209.
- Zsolnay, Á., Baigar, E., Jimenez, M., Steinweg, B., and Saccomandi, F. (1999). Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. *Chemosphere*, 38(1):45–50.
- Zsolnay, Á. and Steindl, H. (1991). Geovariability and biodegradability of the water-extractable organic material in an agricultural soil. *Soil Biology and Biochemistry*, 23(11):1077–1082.